"Los programas dirigidos al mejoramiento del bienestar humano en las zonas de conflicto y transición del mundo en desarrollo se enfrentan a retos sin precedentes y por esto que es tan importante y oportuno el libro de Ray Rist y Linda Morra. Con un gran acopio de conocimientos y experiencia excepcional, el libro organiza en un formato conveniente los fundamentos de las buenas prácticas de la evaluación en el campo del desarrollo. Por su tratamiento autorizado y bien documentado de los problemas de evaluación en el mundo real, que ha creado el complejo y volátil entorno operativo para el desarrollo, debe ocupar un lugar en la biblioteca de todos los profesionales de la evaluación y el desarrollo".

Robert Picciotto, ex *Director General de Evaluación del Grupo del Banco Mundial; profesor visitante en el King's College de Londres*

"En este libro se ejemplifica todo lo que espero encontrar en un texto sobre evaluación: material bien probado, cuidadosamente evaluado y demostrado en la práctica. Verosímil por cuanto ha sido realizado por personas que saben de lo que hablan por su experiencia y pericia. Contenido que enseña y ejemplifica los estándares de la profesión, y es así tanto útil y práctico como ético y preciso. Hace de la evaluación algo comprensible, accesible y factible, y comunica la importancia y el valor del control y la evaluación. Constituye una buena mezcla de la teoría con la práctica, metodológicamente sólida, ilustrada con ejemplos aplicables y comprensibles. Material del mundo real y realista, es decir basado en las realidades políticas, económicas, culturales y administrativas del desarrollo. Una narrativa que enseña e inspira. Todas esas son las cosas que espero encontrar en un libro y éste cumple con todos y cada uno de esos criterios".

Michael Q. Patton, ex *presidente de la American Evaluation Association; autor de* Utilization-Focused Evaluation, and Qualitative Research and Evaluation Methods; *consultor independiente en evaluación y desarrollo organizativo*

"La evaluación del desarrollo está creciendo con rapidez en los países menos avanzados. La estructura y enfoque de este libro hacen que sea útil para una gran variedad de usuarios, como especialistas y profesionales en evaluación, autores de políticas, parlamentarios, funcionarios del sector público y estudiantes de evaluación del desarrollo. Este libro representa una contribución significativa a la evaluación del desarrollo en el mundo en desarrollo".

Kabir Hashim, *miembro del parlamento de Sri Lanka; ex Ministro de Educación Terciaria y Capacitación; consultor en evaluación del desarrollo*

"Nunca ha sido más urgente la necesidad de un libro como éste, cuando la evaluación de la cooperación para el desarrollo es cada vez más importante. Los lectores, a quienes se presenta una cobertura integral y a la vez concisa de los principales temas y de los intereses actuales, adquirirán conocimientos valiosos en este campo de investigación que se hace cada vez más complejo. Constituye una lectura esencial tanto para los principiantes como para los evaluadores del desarrollo más experimentados. Está escrito con claridad y es de fácil comprensión; cuenta con una variedad de ejemplos del mundo real y valiosas joyas de conocimientos prácticos. Con seguridad se convertirá en el texto más indispensable y será ampliamente leído por quienes deseen conocer la evaluación del desarrollo contemporánea".

Penny Hawkins, *Jefe de Evaluación de la New Zealand Agency for International Development; Vicepresidenta de la Red del Comité de Asistencia para el Desarrollo, sobre Evaluación del Desarrollo*

MAYOL EDICIONES

BANCO MUNDIAL

IEG WORLD BANK | IFC | MIGA
INDEPENDENT EVALUATION GROUP

En trámite

Morra Imas
y Rist

EL CAMINO HACIA LOS RESULTADOS
Diseño y realización de evaluaciones eficaces para el desarrollo

MAYOL EDICIONES

Texto interactivo en www.worldbank.org/r2r

EL CAMINO HACIA LOS RESULTADOS

Diseño y realización de evaluaciones eficaces para el desarrollo

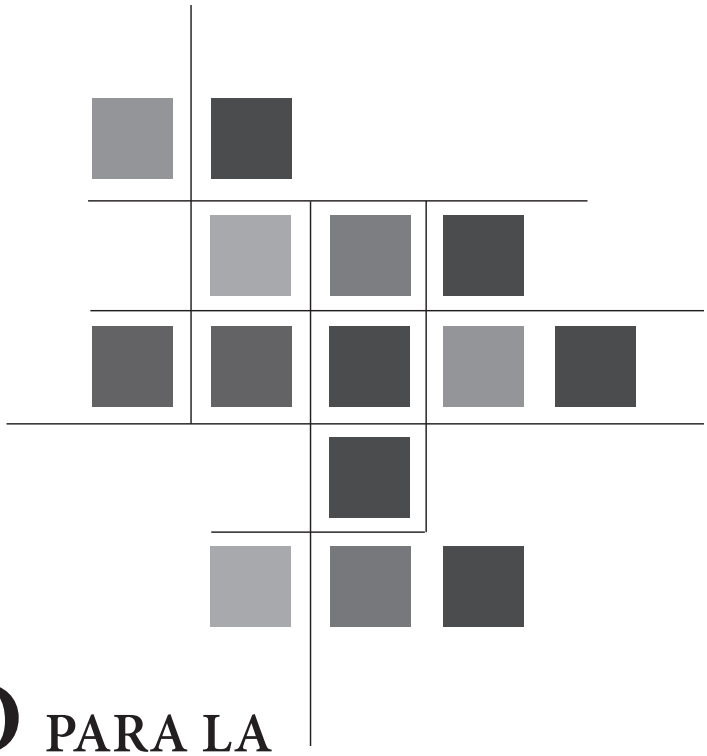
Linda G. Morra Imas
Ray C. Rist

BANCO MUNDIAL

MAYOL EDICIONES

000El camino.indd 1 08/11/2010 03:25:01 p.m.

000El camino.indd 2 08/11/2010 03:25:01 p.m.



EL CAMINO PARA LA OBTENCIÓN DE RESULTADOS

Diseño y realización de evaluaciones eficaces para el desarrollo

Linda G. Morra Imas
Ray C. Rist

BANCO MUNDIAL

MAYOL EDICIONES

000El camino.indd 3 08/11/2010 03:25:02 p.m.

The findings, interpretations, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent.
The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this volume do not imply any judgment concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Los hallazgos, interpretaciones y conclusiones expresados aquí son los del (los) autor (es) y no reflejan necesariamente las opiniones de los directores del Banco Mundial, o de los gobiernos que ellos representan.
El Banco Mundial no garantiza la exactitud de los datos incluidos en este trabajo. Las fronteras, los colores, los nombres y otra información expuesta en cualquier mapa de este volumen no denotan, por parte del Banco, juicio alguno sobre la condición jurídica de ninguno de los territorios, ni aprobación o aceptación de tales fronteras.

This work was originally published by The World Bank in English as *The Road to results. Designing and Conducting Effective Development Evaluations* in 2009. This Spanish translation was arranged by Mayol Ediciones. Mayol Ediciones is responsible for the quality of the translation. In case of any discrepancies, the original language will govern.

Publicado originalmente en inglés como: *The Road to results. Designing and Conducting Effective Development Evaluations* por el Banco Mundial en 2009. La traducción al castellano fue hecha por Mayol Ediciones, editorial que es responsable de su precisión. En caso de discrepancias, prima el idioma original.

© 2009 The International Bank for Reconstruction and Development/The World Bank
© 2009 Banco Internacional de Reconstrucción y Fomento/Banco Mundial
1818 H Street, NW
Washington, DC 20433, USA

Todos los derechos reservados

Primera edición en castellano: noviembre de 2010
Para esta edición:
© 2010 Banco Mundial y Mayol Ediciones S.A.
www.mayolediciones.com

ISBN 978-958-8307- ??? (En trámite)

Diseño de cubierta: Naylor Design
Traducción al castellano: Ignacio Caviedes
Coordinación editorial: María Teresa Barajas S.
Edición y diagramación: Mayol Ediciones S.A.

Impreso y hecho en Colombia - Printed and made in Colombia

Dedicatoria

A mi esposo Silvio, quien me apoya y estimula en mis esfuerzos profesionales, y con especial gratitud a dos mujeres fuertes que han sido mis modelos en evaluación y mis mentoras:
Eleanor Chelimsky y Lois-ellin Datta

LINDA G. MORRA IMAS

Para Molly, Ethan, Madeleine, Lukas, Nina y Anna:
mis seis nietos y nietas que cada día le dan gozo a mi vida

RAY C. RIST

El camino hacia la obtención de resultados está disponible como un texto interactivo en **http://www.worldbank.org/r2r.**

Esta herramienta hace posible que estudiantes y profesores compartan notas y materiales relacionados para mejorar su aprendizaje multimedia.

000El camino.indd 6 08/11/2010 03:25:02 p.m.

Contenido

Prefacio xvii

Acerca de los autores xix

Siglas y abreviaturas xxi

Introducción xxv

FUNDAMENTOS 1

Capítulo 1. Presentación de la evaluación para el desarrollo 3
¿Qué es la evaluación? 3
Orígenes e historia de la disciplina de evaluación 14
El contexto de la evaluación para el desarrollo 21
Principios y estándares para la evaluación para el desarrollo 24
Ejemplos de evaluaciones de desarrollo 31
Resumen 35
Actividad del capítulo 1 36
Referencias y lecturas adicionales 36
Sitios web 39

Capítulo 2. Conocimiento de los temas que orientan la evaluación para el desarrollo 43
Visión general de la evaluación en países desarrollados y en desarrollo 43
Implicaciones de los problemas emergentes para el desarrollo 63
Resumen 85
Referencias y lecturas adicionales 86
Sitios web 91

PREPARAR Y REALIZAR EVALUACIONES EFICACES PARA EL DESARROLLO 93

Capítulo 3. Construir un sistema de seguimiento y evaluación basado en resultados 95
Importancia del seguimiento y la evaluación basados en resultados 95

¿Qué son seguimiento y evaluación basados en resultados? 97
Seguimiento y evaluación tradicionales frente a basados en resultados 97
Diez pasos para construir un sistema de seguimiento y evaluación basado en resultados 102
Resumen 123
Actividades del capítulo 3 124
Referencias y lecturas adicionales 127
Sitios web 127

Capítulo 4. Conocer el contexto de la evaluación y la teoría programática del cambio 129
Análisis del preliminar 129
Identificar el cliente principal y las partes interesadas clave 132
Conocer el contexto 135
Acudir al conocimiento existente 137
Construir, utilizar y evaluar una teoría del cambio 138
Terminología y representaciones gráficas de la teoría del cambio 151
Evaluar una teoría del cambio 156
Resumen 158
Actividades del capítulo 4 159
Referencias y lecturas adicionales 161
Sitios web 163

Capítulo 5. Consideración sobre el enfoque de la evaluación 167
Enfoques generales en la evaluación 167
Resumen 192
Referencias y lecturas adicionales 198
Actividad del capítulo 5 198
Sitios web 201

DISEÑO Y REALIZACIÓN 205

Capítulo 6. Desarrollo de preguntas de la evaluación e inicio de la matriz de diseño 207
Fuentes de preguntas 207
Tipos de preguntas 208
Identificar y seleccionar las preguntas 215
Desarrollo de preguntas buenas 217
Diseño de la evaluación 219
Resumen 226
Actividades del capítulo 6 228
Referencias y lecturas adicionales 229
Sitios web 229

Capítulo 7. Selección de diseño para preguntas de evaluación de causa y efecto, normativas y descriptivas **231**
Conectar las preguntas al diseño 231
Diseños con preguntas de causa y efecto 235
Diseño con preguntas descriptivas 256
Necesidad de diseños de evaluación más rigurosos 259
Actividades del capítulo 7 265
Referencias y lecturas adicionales 268
Sitios web 270

Capítulo 8. Seleccionar y construir instrumentos de recolección de datos **271**
Estrategias de recolección de datos 271
Enfoque estructurado 272
Características de las buenas medidas 274
Datos cuantitativos y cualitativos 276
Resumen 328
Actividades del capítulo 8 329
Referencias y lecturas adicionales 330
Sitios web 333

Capítulo 9. Seleccionar la estrategia de muestreo **335**
Introducción al muestreo 335
Tipos de muestras: aleatorias y no aleatorias 336
Establecer el tamaño de la muestra 344
Resumen 346
Actividades del capítulo 9 348
Referencias y lecturas adicionales 349
Sitios web 349

Capítulo 10. Planear y hacer el análisis de datos **351**
Estrategia de análisis de datos 351
Análisis de datos cualitativos 352
Análisis de datos cuantitativos 364
Vínculo de datos cuantitativos y cualitativos 378
Resumen 381
Actividades del capítulo 10 382
Referencias y lecturas adicionales 384
Sitios web 386

AFRONTAR LOS DESAFÍOS **389**

Capítulo 11. Evaluación de intervenciones complejas **391**
Visión general de la evaluación para el desarrollo 391

Evaluaciones conjuntas 393
Evaluaciones de programas de países 395
Evaluaciones de programas sectoriales 399
Evaluaciones temáticas 400
Evaluación de programas asociativos globales y regionales 404
Resumen 409
Referencias y lecturas adicionales 409
Sitios web 411

DIRIGIR 413

Capítulo 12. Administrar una evaluación 415
Manejo de la matriz de diseño 415
Contratar la evaluación 416
Papeles y responsabilidades de los distintos actores 419
Dirección de personas, tareas y presupuestos 425
Resumen 430
Actividades del capítulo 12 433
Referencias y lecturas adicionales 435
Sitios web 437
Ejemplos de términos de referencia 437

Capítulo 13. Presentación de resultados 439
Ensamblar una estrategia de comunicación 439
Redactar informes de evaluación 441
Presentar información en forma visual 446
Hacer una presentación oral 459
Resumen 462
Actividades del capítulo 13 463
Referencias y lecturas adicionales 463
Sitios web 464

Capítulo 14. Guiar al evaluador: ética, política, estándares y principios orientadores 465
Comportamiento ético 465
Política y evaluación 469
Estándares de evaluación y principios orientadores 476
Ética de evaluación para el sistema de la ONU 478
Resumen 480
Actividad del capítulo 14 481
Referencias y lecturas adicionales 482
Sitios web 483

Capítulo 15. Mirando hacia el futuro **485**
Del pasado al presente 485
El futuro 487
Actividad del capítulo 15 496
Referencias y lecturas adicionales 497
Sitios web 498

Apéndices **499**

Cuadros

1.1 Ejemplos de evaluaciones de políticas, programas y proyectos 10
1.2 Comparación de seguimiento y evaluación 12
1.3 El cambio en el enfoque del Banco Mundial sobre el desarrollo, 1950-2000 23
1.4 Criterios e indicadores para establecer la independencia de las organizaciones de evaluación 28

3.1 Principales componentes de una teoría del cambio 99
3.2 Matriz para la selección de indicadores 107
3.3 Ejemplo del marco de desempeño: resultados e indicadores 107
3.4 Ejemplo de marco de desempeño: resultados, indicadores y datos básicos 110
3.5 Ejemplo de marco de desempeño: resultados, indicadores, datos básicos y objetivos 112
3.6 Cuadro de ejemplo de presentación de resultados 118

4.1 Los papeles de las partes interesadas en una evaluación 134
4.2 Preguntas para hacer en las distintas etapas de una evaluación 137
4.3 Marco lógico para el programa de reducción de la mortalidad infantil 157

5.1 Tipos de preguntas de enfoque al futuro y formas de tratarlas 169
5.2 Características de las técnicas de evaluación participativa y tradicional 180
5.3 Propósito, características, fortalezas y problemas de los distintos enfoques de evaluación 193

6.1 Preguntas cualitativas para plantear durante diferentes fases del ciclo de política 211
6.2 Matriz para seleccionar y clasificar las preguntas de la evaluación 217

7.1 Dos características clave de los diseños experimentales, cuasiexperimentales y no experimentales 235
7.2 Cuatro tipos de auditoría de desempeño 259
7.3 Ventajas y desventajas de diseños experimentales comunes 263

8.1 Muestra de tabla de decisión para método de recolección de datos 272
8.2 Cuándo utilizar un enfoque cuantitativo y cualitativo 280
8.3 Ventajas y problemas de las reuniones comunitarias 283
8.4 Ventajas y problemas de utilizar datos disponibles 291
8.5 Razones para utilizar y ejemplos de observación 292

8.6 Usos de la observación semiestructurada 296
8.7 Ventajas y problemas de la recolección de datos de observación 297
8.8 Ventajas de las encuestas estructuradas y semiestructuradas 299
8.9 Ventajas y problemas de las encuestas 301
8.10 Niveles de refinamiento para determinar el contenido del cuestionario 302
8.11 Resultados de tarjetas de informe de ciudadanos sobre satisfacción con servicios de infraestructura 311
8.12 Usos de las tarjetas de informe de ciudadanos 311
8.13 Ventajas y problemas de las tarjetas de informe de ciudadanos para la recolección de datos 312
8.14 Ventajas y problemas de las entrevistas 314
8.15 Cuatro fases de acciones de grupo focal y moderador 320
8.16 Ventajas y problemas de los grupos focales 321
8.17 Lineamientos para utilizar diarios o bitácoras 322
8.18 Ventajas y problemas de los diarios y las listas de verificación 323
8.19 Ventajas y problemas del juicio de expertos 324
8.20 Ventajas y problemas de la técnica Delphi 327

9.1 Tamaño mínimo requerido de la muestra al nivel de confianza del 95% con un margen de error de 5 puntos porcentuales 345
9.2 Tamaños de muestra para poblaciones de 1 millón y más 346

10.1 Consejos para recopilar y analizar datos cualitativos 353
10.2 Fortalezas y problemas potenciales del análisis de contenido 360
10.3 Organizar e interpretar datos cualitativos 365
10.4 Porcentaje de población urbana por países de América Latina y América Central 369
10.5 Medidas de tendencia central preferidas según el tipo de datos 370
10.6 Opiniones de clientes sobre servicios de salud en la clínica local (1) 374
10.7 Opiniones de clientes sobre servicios de salud en la clínica local (2) 375

11.1 Seleccionar, organizar y administrar un equipo de evaluación externa e informar los efectos directos 396

12.1 Muestra de mapa de tareas 428
12.2 Las 20 actividades de Greer para un administrador de proyecto 431

13.1 Ejemplo de lista de verificación sobre la estrategia de comunicación para las fases preliminar y de diseño de la evaluación 441
13.2 Sistema de seguimiento de recomendaciones 446
13.3 Partes de los gráficos 451
13.4 Tipo y propósito de los distintos tipos de gráficos 453
13.5 Ejemplo de datos en una tabla con muchas líneas 457
13.6 Ejemplo de datos en una tabla con pocas líneas 457
13.7 Ejemplo de tabla de clasificación 458

Gráficos

3.1 Teoría programática del cambio (modelo lógico) para obtener efectos directos e impactos 100

3.2	Ejemplo de teoría programática del cambio (modelo lógico) para reducir la morbilidad infantil mediante el uso de terapia de rehidratación oral	101
3.3	Diez pasos para diseñar, construir y mantener un sistema de seguimiento y evaluación basado en resultados	102
3.4	Espectro de los métodos de recopilación de datos	109
3.5	Identificar el nivel esperado o deseado de mejoramiento requiere la selección de objetivos de desempeño	112
3.6	Tipos clave de control	113
3.7	Vínculos entre el control de la ejecución y el de los resultados	114
3.8	Ejemplo de vínculo entre el control de ejecución y el de los resultados	114
3.9	El logro de resultados por medio de la asociación	115
4.1	Paso de insumos a resultados	140
4.2	Influencias ambientales potenciales en los resultados del programa	141
4.3	Proceso de construir una teoría del cambio	142
4.4	Una teoría del cambio simple para mejorar la toma de decisiones capacitando evaluadores	145
4.5	Diagrama simple de la teoría del cambio con supuestos básicos identificados	147
4.6	Modelo para la teoría del cambio	147
4.7	Teoría del cambio para un programa de mejoramiento del rendimiento estudiantil en lectura, mediante visitas de los maestros al hogar del estudiante	150
4.8	Teoría del cambio para un programa de uso de talleres de participación para reducir la corrupción en el gobierno	152
4.9	Diagrama de flujo estándar	153
4.10	Diagrama de flujo con énfasis en los supuestos	154
4.11	Cadena de resultados estándar	155
4.A	Teoría del cambio simple para un programa de micropréstamos	159
4.B	Teoría del cambio más compleja para un programa de micropréstamos	160
6.1	Uso de un modelo lógico para enmarcar las preguntas de la evaluación	209
6.2	Enfoque a la evaluación para el desarrollo	225
6.3	Plantilla de la matriz de diseño	227
8.1	Muestra de instrumento de recolección de datos	289
8.2	Muestra de formulario de observación basado en temas	293
8.3	Muestra de formulario de observación estructurada	294
10.1	Diagrama de flujo para investigación de análisis de contenido	359
10.2	Hoja de trabajo en blanco para análisis de datos cualitativos	361
10.3	Hoja de trabajo llena para análisis de datos cualitativos	363
10.4	La distribución normal	371
10.5	Distribuciones no normales	371
10.6	Desviaciones estándar en una distribución normal	372
12.1	Ejemplo de diagrama de Gantt	429
13.1	El MATR de dos etapas de la CFI para el seguimiento de las recomendaciones	447
13.2	Ejemplo de diagrama organizativo	450
13.3	Ejemplo de diagrama de Gantt	451

13.4 Reducciones en la temperatura promedio en la provincia A, febrero-junio, 2008 452
13.5 Los puntajes de lectura difieren en tres colegios del distrito, 2004-2005 453
13.6 Los puntajes en la prueba sobre cuidado de niños varían según la región 454
13.7 Las respuestas a las preguntas son diferentes para asistentes y directores 454
13.8 El tercer trimestre presenta los mayores costos de electricidad en dólares 455
13.9 Comparación de puntajes de prueba y nivel de grado con puntajes promedio 456
13.10 Ejemplo de gráfico innecesariamente confuso 460
13.11 Ejemplo de gráfico estilizado que maximiza la tinta de datos 460

Ilustración

13.1 Impresión de un niño de la aldea antes y después de una intervención 449

Recuadros

1.1 Usos de la evaluación 11
1.2 El Institute of Internal Auditors y la International Organization of Supreme Audit Institutions 19
1.3 Las 10 partes de los estándares de calidad de evaluación del CAD 27
1.4 Pertinencia: la World Food Programme's Evaluation of Food Aid for Relief and Recovery en Somalia 31
1.5 Eficacia: evaluación del apoyo del DfID para los World Food Programme's Efforts en Bangladesh 32
1.6 Eficiencia: evaluación de los Disasters Emergency Committee's Mozambique Flood Appeal Funds 33
1.7 Impacto: Joint Evaluation of Emergency Assistance to Rwanda 34
1.8 Sostenibilidad: evaluación de la JICA del Third Country Training Program on Information and Communication Technology 35

2.1 Los Objetivos de Desarrollo del Milenio 67

3.1 El poder de medir los resultados 97
3.2 Diferencia entre control basado en resultados y evaluación basada en resultados 98
3.3 Diez usos de los resultados obtenidos 120

4.1 Cómo efectuar un análisis de partes interesadas 136
4.2 Acudir al fondo de conocimientos sobre la prevención del crimen 138
4.3 Revisar la literatura sobre programas para mejorar el logro estudiantil en un país de África meridional 144

5.1 Evaluación social del Proyecto de Desarrollo y Crédito Agrícola de Azerbaiyán 177
5.2 Generar confianza a través de la evaluación participativa 183
5.3 Uso de una síntesis para medir los beneficios ambientales 188

6.1 Evaluar una intervención que no tiene estándares 212

6.2	Evaluar políticas e intervenciones utilizando preguntas y respuestas	214
6.3	Las cinco etapas del proceso de la evaluación	223
7.1	¿Qué hace irse a los elefantes?	232
7.2	¿Funcionan los colegios administrados por la comunidad?	246
7.3	Uso de un diseño de corte transversal para responder preguntas descriptivas	252
7.4	Ejemplo de diseño de caso de estudio para preguntas descriptivas	256
7.5	Uso de diseño de antes y después para responder preguntas descriptivas	257
7.6	Ejemplo de diseño de series de tiempo interrumpidas para preguntas descriptivas	258
7.7	Impacto de programas de capacitación de empleo para trabajadores despedidos	260
7.8	El apoyo de España a los diseños de evaluación rigurosos	261
8.1	Reglas para la recolección de datos	273
8.2	Adoptar un enfoque estructurado para evaluar una intervención agrícola	274
8.3	La lista de 20 preguntas de revisión cualitativa de Patton	278
8.4	Medir la popularidad de exposiciones de arte, utilizando métodos inoportunos y discretos	281
8.5	Uso de Google Earth para la cartografía	285
8.6	Preguntas de encuestas estructuradas y semiestructuradas	299
8.7	Consejos para redactar preguntas de encuestas	303
8.8	Lineamientos generales para efectuar encuestas	309
8.9	Consejos para escribir cuestionarios	309
8.10	Consejos para realizar entrevistas	313
8.11	Consejos para realizar entrevistas en diferentes culturas	314
8.12	Consejos para diseñar preguntas de grupos focales	319
8.13	Muestra de preguntas de grupo focal	320
9.1	Usar muestreo de grupos para identificar personas con Sida y entrevistarlas	341
10.1	Ejemplo de codificación	356
10.2	Uso del análisis de contenido para identificar los beneficios percibidos de las bombas manuales	357
10.3	Estrategias para la codificación electrónica de datos	360
10.4	Consejos para mantener datos cuantitativos	366
10.5	Cálculo de la desviación estándar	373
10.6	Lineamientos para analizar datos cuantitativos de encuestas	374
11.1	Ejemplo de metodología para evaluación del programa de país	398
11.2	Evaluación conjunta externa del sector de salud de Tanzania, 1999-2006	401
11.3	Ejemplo de evaluación temática: evaluación sobre niños que buscan en la basura en África, Asia y Europa	403
12.1	Consejos para escribir los términos de referencia para una evaluación	419
12.2	Consejos para resolver conflictos	426
12.3	Consejos para mejorar el equipo de trabajo	427
13.1	Uso de técnicas innovadoras de comunicación para aumentar el interés en los resultados de la evaluación	442

13.2 Consejos para crear gráficos eficaces 459
13.3 Consejos para preparar presentaciones en PowerPoint y volantes 461
13.4 Consejos para el uso de diapositivas 462

15.1 Temas básicos para diagnosticar los sistemas de SyE de un gobierno 490
15.2 Competencias esenciales para evaluadores de programas 493

Prefacio

La evaluación para el desarrollo es un componente dinámico dentro del mosaico general de las evaluaciones, y aborda los campos profundos, complejos y emergentes del desarrollo. Las evaluaciones sobre el alivio de la pobreza, la globalización y su impacto en los pobres, las consecuencias del calentamiento global en los países menos favorecidos, las desigualdades estructurales de los sistemas financieros mundiales y las estrategias de ayuda a países con recientes conflictos, son sólo unos pocos campos en donde la evaluación para el desarrollo contribuye, en realidad, a nuestro conocimiento de la respuesta que damos a estos temas apremiantes.

A medida que aumenta la presión en todo el mundo por la rendición de cuentas de los gobiernos y las organizaciones por las consecuencias de sus acciones, en pos de un mayor grado de reacción ante partes interesadas internas y externas por su desempeño, y de manera más profunda, por una mayor eficacia en los esfuerzos para el desarrollo, surge la evaluación como una forma clave para abordar sistemáticamente y responder la pregunta: "¿Y qué?" No basta con documentar que se está ocupado; ahora es un requisito documentar que se es (o no) eficaz.

La evaluación para el desarrollo constituye también un nuevo campo de investigación. Encontrar formas para evaluar es algo insustancial, cuando son débiles o inexistentes los sistemas de datos gubernamentales, frecuente la corrupción de información para fines políticos, grandes y reales los vacíos de información, y no se cuenta con la seguridad de que la información provista sea confiable. Ante estos desafíos, la evaluación para el desarrollo es elástica, innovadora y creativa para encontrar maneras de suministrar información para los ciudadanos, funcionarios oficiales, donantes, sociedad civil y los medios, en cuanto a si los programas oficiales están marcando la diferencia.

Con este texto se busca contribuir al fortalecimiento de la evaluación para el desarrollo como herramienta para que la creación y la implementación de políticas y programas, en lo particular, y de sistemas de gobernanza, en lo general, se sustenten en información sólida. La evaluación puede constituir una poderosa herramienta de administración pública para mejorar la forma como operen los gobiernos y las organizaciones, y obtengan resultados. Su contribución no termina en el umbral del sector público. La evaluación puede constituir una poderosa herramienta para la sociedad

000El camino.indd 17 08/11/2010 03:25:02 p.m.

civil, las organizaciones no gubernamentales (ONG) y las organizaciones donantes que buscan darle soporte al desarrollo de los pobres.

El material de este libro es una adaptación del Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET), que es un programa que ofrecen el Banco Mundial y la Carleton University cada verano (*ver* www.IPDET.org). El IPDET reúne participantes de todo el mundo para dedicar un mes al estudio de la evaluación para el desarrollo. El material de este libro constituye una elaboración del curso que ofrece el IPDET, y se suministra con el propósito de expandir y compartir el contenido del curso con otros interesados en la evaluación para el desarrollo.

Nosotros, los autores, tenemos una deuda de gratitud con varias personas que aportaron su consejo, leyeron partes del manuscrito y ofrecieron críticas, y nos alentaron a continuar para hacer de este libro una realidad. La lista completa de las personas a quienes deseamos agradecer se encuentra al final del libro, en el apéndice 1. Sin embargo, aquí agradecemos sus esfuerzos a un grupo selecto de personas, a saber: Michael Patton, Patrick Grasso, Martin Abrams, Niels Dabelstein, Gregg Jackson, Gene Swimmer y Nancy Porteous; cada uno de ellos leyó y criticó secciones del libro. Santiago Pombo Bejarano, de la Oficina del editor del Banco Mundial, fue el impulso y su estímulo nunca flaqueó para continuar el trabajo en este manuscrito.

También agradecemos a dos personas que fueron nuestros asociados en la empresa de hacer que el IPDET fuese una realidad en la Carleton University: Karen Gisnberg y Barbara Levine. No habría libro de no existir el IPDET, y ellas son unas socias y amigas maravillosas.

Por último, debemos reconocer la destacada contribución de Diane Schulz Novak, quien trabajó con nosotros a lo largo de todo el proceso de escribir una y otra vez el manuscrito. Su dedicación, atención y arte trabajando a nuestro lado fueron tan esenciales que en su ausencia no estaríamos escribiendo este prefacio, pues tampoco habría un libro.

Nosotros, Linda y Ray, hemos sido amigos y colegas por cerca de 30 años. Nos conocimos y empezamos a trabajar juntos en la U.S. Government Accountability Office, en 1981. Hemos estado juntos en el Banco Mundial ya por más de una década y la colaboración y la amistad se han hecho más y más fuertes. Es justo que cuando llegamos a la cima de nuestras carreras podamos ofrecer a la comunidad de evaluación este fruto de nuestro trabajo conjunto.

Acerca de los autores

Linda G. Morra Imas es asesora en el Independent Evaluation Group del Grupo del Banco Mundial; es conocida como "la madre del IPDET", el Programa Internacional para la Capacitación en Evaluación para el Desarrollo, del cual es codirectora. Asesora y entrenadora en control y evaluación en muchos países del mundo; ha sido profesora adjunta en George Washington University y en Carleton University. Llegó al Grupo del Banco Mundial en 1996 y dirigió numerosas evaluaciones. Fue directora superior en la United States Government Accountability Office, y ha testificado con frecuencia en Comités del Congreso sobre programas de educación y empleo.

Ray C. Rist es asesor en el Independent Evaluation Group del Grupo del Banco Mundial, al que llegó en 1997; su carrera incluye 15 años en el gobierno de los Estados Unidos, con cargos en las ramas ejecutiva y legislativa del poder. Ha ocupado cargos académicos en la Cornell University, The Johns Hopkins University y The George Washington University. Fue Senior Fullbright Fellow en el Instituo Max Planck en Berlín, Alemania, en 1976-77. Es autor, editor o coeditor de 25 libros, ha escrito 135 artículos y dictado conferencias en más de 75 países.

000El camino.indd 19 08/11/2010 03:25:03 p.m.

000El camino.indd 20 08/11/2010 03:25:03 p.m.

Siglas y abreviaturas

ACAC	Análisis cualitativo asistido por computador
AEA	American Evaluation Association
ANOVA*	análisis de variancia
ANSI	American National Standards Institute
AOD	ayuda oficial para el desarrollo
ASEAN	Association of Southeast Asian Nations
BAfD	Banco Africano de Desarrollo
BAsD	Banco Asiático de Desarrollo
BERD	Banco Europeo para la Reconstrucción y el Desarrollo
BMD*	Banco(s) Multilateral(es) de Desarrollo
CAD	Comité de Asistencia para el Desarrollo
CAHMI	Child and Adolescent Health Measurement Initiative
CEI	Comunidad de Estados Independientes
CES	Canadian Evaluation Society
CESE*	Europa central, oriental y meridional
CFI	Corporación Financiera Internacional
CGD*	Centro para el Desarrollo Mundial
CIDA	Canadian International Development Agency
CIID	Centro Internacional de Investigaciones para el Desarrollo
CODE	Committee on Development Effectiveness
CSF	Community Support Framework
DANIDA	Danish International Development Agency
DEC	Disasters Emergency Committee
DfID	Department for International Development
ECI	Early Communication Indicator
EDUCO	Programa de Escuelas Manejadas por la Comunidad de El Salvador
EET*	gravado-exento-exento
EN	Entorno en los negocios
ES	enfoque(s) sectorial(es)
ESEA	Elementary and Secondary Education Act
ESRC	Economic and Social Research Center

000El camino.indd 21 08/11/2010 03:25:03 p.m.

ETAL*	Encuestas telefónicas asistidas por computador
FAO*	Organización de las Naciones Unidas para la Agricultura y la Alimentación
FIT-ED	Foundation for Information Technology and Education Development
FMI	Fondo Monetario Internacional
FNUAP	Fondo de Población de las Naciones Unidas
GAFI	Grupo de Acción Financiera contra el Blanqueo de Capitales
GAO	Government Accountability Office
GICHD*	Centro Internacional de Desminado Humanitario de Ginebra
GPRA*	Government Performance Results Act
GRPP	Global and Regional Partnership Programs
GyD	género y desarrollo
HSSP2*	Plan de Apoyo del Sector Salud 2
HZZO	Croatian Institute for Health Insurance
IBSTPI*	Junta Internacional de Estándares para la capacitación, desempeño e instrucción
IDEAS	Asociación Internacional de Desarrollo de Evaluación
IED	inversión extranjera directa
IFAD	International Fund for Agricultural Development
IISD	International Institute for Sustainable Development
IMG	ingreso mínimo garantizado
INTOSAI*	Organización Internacional de Entidades Fiscalizadoras Superiores
IOB	Policy and Operations Evaluation Department (Países Bajos)
IOCE*	Organización Internacional para la Cooperación en Evaluación
IPDET*	Programa Internacional para la Capacitación en Evaluación para el Desarrollo
IRBM*	sistema integrado de administración basada en resultados
ISR	interpretación subjetiva de la realidad
JEEAR	Joint Evaluation of Emergency Assistance to Rwanda
JICA*	Agencia Japonesa de Cooperación Internacional
LGAB*	ten buen aspecto-evita la vergüenza
LSMS	Living Standards Measurement Survey Study
MATR	Management Action Tracking Record
MBS	Modified Budgeting System
MFPED*	Ministerio de Finanzas, Planeación y Desarrollo Económico
MIS*	sistema de información para la administración
MMV*	Medicines for Malaria Venture
MSH	Management Sciences for Health
NAD*	Departamento Nacional de Auditoría
NCSTE	Chinese National Centre for Science and Technology Evaluation
NDA	nivel de adopción
NDC*	contribución definida especulativamente
NEIR TEC	Northeast and the Islands Regional Technology in Education Consortium
NHIFA	National Health Insurance Fund Administration (Hungría)

NORAD*	Agencia noruega de Cooperación para el Desarrollo
NSGPR	National Strategy for Growth and Reduction of Poverty
NSSI	National Social Security Institute (Bulgaria)
OCDE	Organización para la Cooperación y el Desarrollo Económicos
ODM	Objetivos de Desarrollo del Milenio
OIM	Organización Internacional para las Migraciones
OIT	Organización Internacional del Trabajo
OMC	Organización Mundial del Comercio
OMS	Organización Mundial de la Salud
ONG	Organización(es) no Gubernamental(es)
PEAP*	Plan de Acción para la Erradicación de la Pobreza
PIB	producto interno bruto
PNUD	Programa de las Naciones Unidas para el Desarrollo
POW*	Programa de Trabajo para el Sector de la Salud
PPAE	Países pobres altamente endeudados
PPBS*	Sistema de Planeación, Programación y Presupuesto
PPBS	Program Performance Budgeting System
PRSP*	Documento(s) de Estrategias de Producción de la Pobreza
RSE	responsabilidad social de las empresas
SDC	Agencia Suiza de Cooperación para el Desarrollo
SHIPDET	Shanghai International Program for Development Evaluation Training
SIDA	Agencia Sueca Internacional de Cooperación para el Desarrollo
SIEF	Spanish-World Bank Trust Fund for Impact Evaluation
SP	solicitud de propuesta(s)
SyE	seguimiento y evaluación
TI	Transparencia Internacional
TR	términos de referencia
UE	Unión Europea
UNEG*	Grupo de Evaluación de las Naciones Unidas
UNEP*	Programa de las Naciones Unidas para la Protección del Medio Ambiente
UNESCO*	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
UNHCR*	Alto Comisionado de las Naciones Unidas para los Refugiados
UNICEF*	Organización de las Naciones Unidas para la Infancia
UNSD*	División de Estadísticas de las Naciones Unidas
USAID*	Agencia de los Estados Unidos para el Desarrollo Internacional
WID*	mujeres en el desarrollo
ZUS	Social Insurance Fund (Polonia)

* Por sus iniciales en inglés.

000El camino.indd 24 08/11/2010 03:25:03 p.m.

Introducción

El marco analítico, conceptual y político para el desarrollo está cambiando en forma drástica. La nueva agenda para el desarrollo requiere un conocimiento más amplio de los sectores, países, estrategias y políticas para el desarrollo y hace énfasis en el aprendizaje y la retroalimentación continua de todas las fases del ciclo de desarrollo.

En realidad, puede considerarse la evaluación para el desarrollo como una especie de bien público:

> La evaluación sobrepasa los límites de cualquier organización individual. Un buen estudio de evaluación puede tener efectos de resultados positivos en toda la comunidad para el desarrollo. La evaluación para el desarrollo tiene las características de un bien público internacional (Picciotto y Rist, 1995, p. 23).

A medida que la agenda para el desarrollo avanza en su alcance y complejidad, la evaluación para el desarrollo le sigue los pasos. Los evaluadores para el desarrollo se alejan de la implementación tradicional y los modelos de evaluación orientados al producto, y se acercan a los modelos de evaluación basados en resultados, pues la comunidad para el desarrollo demanda resultados y acoge los Objetivos de Desarrollo del Milenio (ODM). Al alejar la comunidad para el desarrollo su enfoque en los proyectos a fin de atender integralmente los retos del país, los evaluadores para el desarrollo buscan métodos con los cuales evaluar los resultados a nivel de país, sector, tema, política, y aun el nivel global. Cuando la comunidad para el desarrollo reconoce la importancia de un enfoque no sólo integral, sino también coordinado hacia los retos de los países en desarrollo y hace hincapié en las asociaciones, los evaluadores para el desarrollo se dedican cada vez más a las evaluaciones conjuntas, que, aunque ventajosas en muchos aspectos, agregan complejidad a la evaluación para el desarrollo (OCDE, 2006). Además, los evaluadores para el desarrollo se enfrentan más y más al reto de la medición del desempeño de una organización individual de desarrollo en este contexto más amplio y de identificar su contribución.

Con la llegada de este enfoque más complejo y demandante para el desarrollo, se ha hecho más difícil de diseñar la evaluación, y ahora comprende demandas metodológicas más intrincadas y fija altos estándares para determinar los impactos.

000El camino.indd 25 08/11/2010 03:25:03 p.m.

La demanda de nuevos enfoques de evaluación y una nueva combinación de destrezas va más allá de la economía. Los asuntos urgentes, como el cambio climático, exigen nuevos enfoques para evaluar la sostenibilidad. El alcance de los problemas ambientales, las consecuencias multinacionales, las dificultades para obtener medidas comparables y la evidencia persistente de consecuencias no previstas requieren un enfoque complejo y multimetódico para la evaluación.

> Puede muy bien esperarse que ninguna disciplina individual predomine en un empeño que aborda los múltiples retos, esperanzas y esfuerzos excesivos de la mayoría de la humanidad. En ausencia de un solo punto individual de consolidación, es común traspasar fronteras disciplinarias, y los evaluadores son cada vez más eclécticos y atrevidos en su uso de los instrumentos de las ciencias sociales (Picciotto y Rist, 1995, p. 169).

La formación de la capacidad de evaluación –creando unidades de evaluación capacitadas en la práctica y los métodos de la evaluación para el desarrollo– constituye un reto para la mayor parte de los países en desarrollo. El incremento en las asociaciones nacionales de evaluación, en los países en desarrollo y en los grupos de evaluaciones regionales, constituye un importante primer paso para el mayor profesionalismo de la comunidad de evaluación para el desarrollo. En la todavía joven International Development Evaluation Association (IDEAS), los evaluadores para el desarrollo encuentran ya una organización internacional profesional dedicada a atender sus necesidades. Además, el crecimiento en los cursos a nivel de posgrado universitario y los centros de capacitación regionales contribuyen también a la formación de capacidad para la evaluación para el desarrollo.

El propósito de este texto es que se constituya en una herramienta para la formación de capacidad en evaluación para el desarrollo, y está dirigido a ayudar a los evaluadores para el desarrollo a reflexionar y explorar la nueva arquitectura de la evaluación, y especialmente a diseñar y realizar evaluaciones centradas en los resultados obtenidos al enfrentar los retos para el desarrollo.

El Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET) fue creado por el Operations Evaluation Department (ahora el Independent Evaluation Group, IEG) del Banco Mundial, en 2001. En asociación con la Carleton University, el IEG realizó el primer programa, con apoyo del Instituto del Banco Mundial, en Ottawa, Canadá. Desde 2003, el programa ha venido ofreciendo versiones adaptadas a las necesidades de una y dos semanas del IPDET, que se han realizado en 12 países. En 2007, se fundó el Shanghai International Program for Development Evaluation Training (SHIPDET).

Si bien el IPDET ha evolucionado continuamente a fin de reflejar la naturaleza cambiante para el desarrollo, continúa dirigido en general a aquellos que trabajan, o están por trabajar, en la evaluación para el desarrollo y busca suministrar las herramientas genéricas para la evaluación de intervenciones para el desarrollo (políticas, programas y proyectos) a niveles local, nacional, regional y global. Se dirige a personal de evaluación de agencias bilaterales y multilaterales, gobiernos de países en desarrollo, ONG, e igualmente a parlamentarios y consultores privados.

000El camino.indd 26 08/11/2010 03:25:03 p.m.

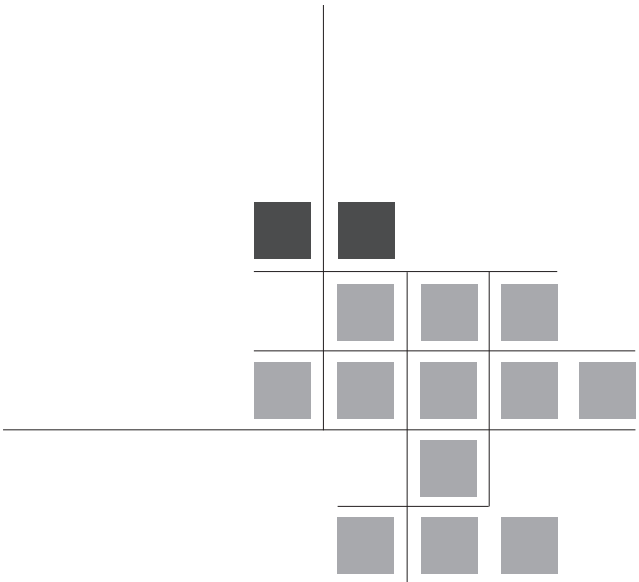
El objetivo general del IPDET es enriquecer el conocimiento, las capacidades y las destrezas para diseñar y realizar evaluaciones para el desarrollo eficaces para una toma de decisiones basada en la evidencia. Se basa en 14 módulos de instrucción que en conjunto constituyen una visión general del camino para la evaluación eficaz de las intervenciones para el desarrollo.

Este volumen se basa en esos módulos y los expande, presentando una discusión integral de los temas que encuentran los evaluadores para el desarrollo, y también es una guía de iniciación a la evaluación para el desarrollo. A través de este texto, muchas más personas de las que trabajan en el desarrollo contarán con herramientas genéricas para producir evaluaciones sólidas de los resultados para el desarrollo.

Referencias

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2006. DAC. *Guidance for Managing Joint Evaluations.* París: OCDE. http://www.oecd.org/dataoecd/28/14/37484787.pdf.

Picciotta, Robert y Ray C. Rist. 1995. *Evaluating Country Development Policies and Programs: New Approaches for a New Agenda.* San Francisco: Jossey-Bass.



FUNDAMENTOS

"El verdadero genio reside en la capacidad de evaluar la información incierta, azarosa y conflictiva."

—WINSTON CHURCHILL

Capítulo 1. Presentación de la evaluación para el desarrollo

- ¿Qué es la evaluación?
- Orígenes e historia de la disciplina de evaluación
- El contexto de la evaluación para el desarrollo
- Principios y estándares para la evaluación para el desarrollo
- Ejemplos de evaluaciones para el desarrollo

Capítulo 2. Conocimiento de los temas que guían la evaluación para el desarrollo

- Visión general de la evaluación en los países desarrollados y en desarrollo
- Implicaciones de las tendencias emergentes en el desarrollo

001El camino.indd 1 08/11/2010 03:27:40 p.m.

001El camino.indd 2 08/11/2010 03:27:40 p.m.

CAPÍTULO 1

Presentación de la evaluación para el desarrollo

En este capítulo se presentan la definición y los conceptos generales fundamentales de la evaluación de proyectos, programas y políticas. Luego se pasa a la evaluación de intervenciones para el desarrollo, llamadas a menudo *evaluación para el desarrollo.*

¿Qué es la evaluación?

Para empezar a comprender la evaluación para el desarrollo, es importante comprender qué se entiende por evaluación, sus propósitos y cómo puede utilizarse. En esta parte del capítulo:

- Se define la evaluación.
- Se identifica el propósito de la evaluación.
- Se relacionan los beneficios de la evaluación.
- Se indica qué evalúan los evaluadores.
- Se describen los usos de la evaluación.
- Se examina la relación entre seguimiento y evaluación.
- Se identifican los papeles y actividades de los evaluadores profesionales.

001El camino.indd 3 08/11/2010 03:27:40 p.m.

Definición de evaluación

Evaluación: Determinación del valor de un proyecto, un programa o una política.

La *evaluación* se ha definido de muchas maneras, y en el Oxford English Dictionary se define como:

> 1. La acción de valorar o avaluar (bienes, etc.); un cálculo o una declaración de valor; 2. La acción de evaluar o de determinar el valor de (una expresión matemática, una cantidad física, etc.) o de estimar la fuerza de las probabilidades, la evidencia, etc.

Dentro de la disciplina de la evaluación, el término ha llegado a tener varios significados, y las diferencias en las definiciones reflejan distintos énfasis en el propósito de la evaluación –rendición de cuentas frente a aprendizaje– o la oportunidad de la evaluación en relación con la madurez del programa, el proyecto o la política. En realidad, no existe un acuerdo general sobre la misma definición.

> De hecho, al considerar el papel del lenguaje en la evaluación, Michael Scriven, uno de los fundadores de la evaluación moderna, anotaba recientemente que existen cerca de sesenta términos distintos para la evaluación, que se aplican a uno u otro contexto. Entre ellos están: *fallar, valorar, analizar, evaluar, criticar, examinar, graduar, inspeccionar, juzgar, tasar, clasificar, revisar, calificar, estudiar, someter a prueba…* (Fitzpatrick, Sanders y Worthen, 2004, p. 5).

La mayoría de las definiciones de evaluación incluye el concepto de hacer un juicio del valor o el mérito del objeto de la evaluación. En realidad, este "valorar" se utiliza para diferenciar la evaluación de las actividades de investigación y seguimiento.

La definición utilizada en este volumen proviene de la Organización para la Cooperación y el Desarrollo Económicos (OCDE)/Comité de Asistencia para el Desarrollo (CAD) Glossary (OCDE, 2000, p. 21):

> Evaluación se refiere al proceso de establecer el valor o el significado de una actividad, política o programa, […] tan sistemático y objetivo como sea posible, de una intervención planeada, en curso o completa.

Las evaluaciones pueden ser formativas, finales o prospectivas:

> Las evaluaciones formativas son evaluaciones dirigidas a mejorar el desempeño [y] muy a menudo se efectúan durante la fase de ejecución de los proyectos o programas. Las evaluaciones formativas pueden también realizarse por otras razones, como cumplimiento, requerimientos legales o como parte de una iniciativa más general de evaluación. En contraste, las evaluaciones finales son estudios efectuados al final de una intervención (o una fase de esa intervención) para establecer el grado al cual se produjeron los efectos directos previstos. La evaluación sumativa tiene el objeto de suministrar información sobre el valor de un programa (OCDE, 2002, pp. 21-22).

Evaluación formativa: De las formas como se ejecuta un programa, una política o un proyecto.

En una *evaluación formativa* se tienen en cuenta las formas como se ejecuta un programa, una política o un proyecto. Examina si la "lógica operativa" asumida corresponde a las operaciones reales, o no, e identifica las consecuencias (in-

mediatas) que produce la (etapa de) ejecución. Esta clase de evaluación se efectúa durante la fase de ejecución de un proyecto o un programa. Las evaluaciones formativas se conocen algunas veces como *evaluaciones de proceso,* ya que se concentran en las operaciones.

Un ejemplo de evaluación formativa que se hizo para el Centro Internacional de Investigaciones para el Desarrollo (CIID) de su iniciativa para el manejo de los recursos naturales en América Latina y el Caribe (denominada Minga) (Adamo, 2003). El objetivo general de la iniciativa Minga fue contribuir a la formación de profesionales, hombres y mujeres, en el manejo de los recursos naturales en Bolivia, Ecuador y Perú.

Un componente de la iniciativa del programa que interesó al CIID fue la incorporación del tratamiento de los géneros a las actividades habituales. Para aprender más sobre esta incorporación al programa, el CIID contrató una evaluación formativa. La metodología para ésta empezó con una revisión de los documentos del programa relativos a actividades de incorporación del tratamiento de los géneros. Los evaluadores revisaron también informes de viaje, a fin de evaluar en qué medida se trataba la consideración del género durante las visitas. Las entrevistas se realizaron con miembros del personal del programa para examinar sus esfuerzos individuales y sus experiencias, y para incorporar el tratamiento del género a su trabajo y las enseñanzas que adquirieron durante dichas actividades.

Un tipo de evaluación formativa es una evaluación de término medio o de mitad de camino. Como su nombre indica, esta evaluación se efectúa aproximadamente hacia la mitad de la ejecución de un proyecto, un programa o un cambio de política. El propósito de una evaluación de mitad de camino es identificar las características del programa que están funcionando bien y las que no. Estas evaluaciones pueden empezar a centrarse en las enseñanzas aprendidas, lo mismo que en la pertinencia, la eficacia y la eficiencia. Las enseñanzas aprendidas son importantes para orientar futuras intervenciones y mejorar las actuales.

Una *evaluación sumativa,* llamada a menudo *evaluación de resultados o de impacto,* se realiza al final de una intervención o en una intervención madura para establecer el grado al cual se obtuvieron los resultados previstos. La evaluación sumativa se realiza para suministrar información sobre el valor y el impacto del programa. Estas evaluaciones incluyen evaluaciones de impacto, investigaciones sobre efectividad en costos, cuasi-experimentos, experimentos al azar y casos de estudio.

Evaluación sumativa: Realizada al final de una intervención para establecer el grado al cual se obtuvieron los resultados deseados.

Un ejemplo de *evaluación sumativa* es la efectuada por el Banco Asiático de Desarrollo (BAsD) para evaluar el Segundo Programa del Sector Financiero en Mongolia (BAsD, 2007). El programa incluyó reformas al sector financiero que incluyeron la restructuración y la transformación del sector de un sistema de un solo banco a un sistema de dos líneas apoyado por el BAsD. Se efectuó una *evaluación sumativa* al terminar la segunda fase del programa. Las evaluaciones finales se utilizan para responder preguntas sobre pertinencia, desempeño, impactos, sostenibilidad, utilidad externa y *enseñanzas aprendidas.*

La diferencia entre una evaluación formativa y una final puede expresarse así:

- Las formativas se centran en la ejecución y el mejoramiento del proyecto, el programa o la política.
- Las finales se centran en los resultados. Capacitan a las personas para tomar decisiones sobre la continuidad, la replicación, el escalamiento gradual o la finalización de un programa, un proyecto o una política.

Usualmente las organizaciones requieren y utilizan ambas clases de evaluaciones en momentos diferentes del ciclo de un proyecto, un programa o una política.

En una *evaluación prospectiva* se evalúan los probables resultados de los proyectos, los programas o las políticas. Es algo similar a una evaluación de la evaluabilidad. Una evaluación de la evaluabilidad responde a las preguntas "¿Vale la pena evaluar este programa o proyecto?" y "¿Se justifican el esfuerzo y los recursos empleados para obtener los beneficios esperados?" En una evaluación prospectiva se sintetizan los hallazgos de evaluación de estudios anteriores para evaluar los probables resultados de nuevos proyectos, programas o políticas propuestos. Por ejemplo, los comités del Congreso en los Estados Unidos solicitan con frecuencia asesoría a la Government Accountability Office (GAO) para proyectar los probables resultados de la legislación propuesta.[1]

Evaluación prospectiva: Evalúa los probables resultados de un proyecto, un programa o una política propuesta.

Un ejemplo de evaluación prospectiva, algo antiguo pero interesante de todas formas, es el estudio de 1986 de la GAO *Teenage Pregnancy: 500.000 Births a Year but Few Tested Programs* (U.S. GAO, 1986). En esta evaluación se utilizaron cuatro procedimientos. Se analizaron las principales características de dos proyectos de ley del Congreso, se revisaron estadísticas sobre la magnitud de los embarazos en adolescentes, se examinaron las características de los programas federales y no federales, y se revisaron estudios de evaluación sobre la eficacia de programas anteriores para asistir a las adolescentes embarazadas y con cuidado de hijos, como también a las que presentaban riesgo de embarazo. Los evaluadores reconstruyeron la teoría fundamental del programa y la lógica operativa de ambos proyectos de ley, a fin de averiguar por qué se consideraba que estas iniciativas funcionarían según se proponía en la legislación. Luego compararon la evidencia hallada con las características de la legislación propuesta.

A veces, esta clase de evaluación prospectiva es denominada *evaluación ex ante* (*antes del hecho*) (Rossi y Freeman, 1993). Las evaluaciones prospectivas o *ex ante* incluyen con mucha frecuencia la reconstrucción o la valoración de la teoría del programa y los estudios de escenarios, y también resúmenes de la investigación y evaluación existente, a fin de asegurar el apoyo empírico para las iniciativas propuestas.

---

1 La U.S. General Accounting Office cambió su nombre por el de Government Accountability Office en julio de 2004 y todavía mantiene la siglas GAO.

Propósito de la evaluación

La evaluación puede utilizarse con varios propósitos. Dentro de la disciplina, existen puntos de vista distintos sobre cuál debería ser el propósito o la meta de la evaluación en un contexto dado.

Una opinión predominante es la de que la evaluación tiene cuatro propósitos distintos:

- *Propósito ético:* informar a dirigentes políticos y ciudadanos sobre la ejecución de una política o un programa y de los resultados obtenidos. En este propósito se combinan los objetivos de utilizar una mejor rendición de cuentas, procesar información y servicio a la democracia.
- *Propósito empresarial:* lograr una distribución más racional de los recursos financieros y humanos entre programas "competidores", mejorar la administración de los programas e incrementar sus beneficios.
- *Propósito de decisión:* allanar el camino para las decisiones sobre la continuación, la terminación o la reconfiguración de una política o un programa.
- *Propósito educativo y motivador:* educar y motivar a las agencias públicas y sus asociados, capacitándolos para comprender los procesos en los cuales están vinculadas, y a identificarse con sus objetivos (Scientific and National Councils of Evaluation, 1999).

Los evaluadores destacados en el campo describen los propósitos de la evaluación siguientes:

- Obtener un mejoramiento social.
- Promover y fomentar la democracia deliberativa.
- Aportar supervisión y cumplimiento.
- Asegurar rendición de cuentas y transparencia.
- Generar, compartir y manejar conocimiento.
- Contribuir al mejoramiento organizativo.
- Promover el diálogo y la cooperación entre las partes interesadas clave.
- Determinar la pertinencia, implementación, la eficiencia, la eficacia, el impacto y la sostenibilidad de proyectos, programas o políticas.
- Generar enseñanzas aprendidas.

Chelimsky y Shadish (1997) adoptan una perspectiva global, extendiendo el contexto de la evaluación a los retos mundiales. Los retos que mencionan incluyen el impacto de nuevas tecnologías, desequilibrios demográficos entre las naciones, protección ambiental, desarrollo sostenible, terrorismo, derechos humanos y otros problemas que se extienden más allá de un programa o incluso de un país.

En último término, el propósito de cualquier evaluación es suministrar información a quienes toman las decisiones para que sus decisiones sobre proyectos, programas o políticas sean mejores. La evaluación debería ayudarlos a conocer qué es probable

que suceda, esté sucediendo, o ha sucedido como consecuencia de una intervención, e identificar formas de obtener más de los beneficios deseados.

Beneficios de la evaluación

La evaluación ayuda a responder preguntas sobre las intervenciones, como las siguientes:

- ¿Cuáles son los efectos de la intervención?
- ¿Funciona la intervención como se planeó?
- ¿Existen diferencias en los sitios en cuanto al funcionamiento de la intervención?
- ¿Quién se beneficia con esta intervención?

Las personas se benefician con las intervenciones en formas distintas. Algunas en forma directa, otras en forma indirecta; estas últimas son personas que no participan en la intervención, pero de todas maneras obtienen beneficios de ella. Algunas intervenciones producen beneficios a corto plazo, otras lo hacen a largo plazo.

Para ilustrar las distintas formas como se benefician las personas, considérese el ejemplo siguiente. El U.S. Department of Housing and Urban Development (1997) evaluó un programa de basquetbol de medianoche para muchachos y muchachas de 16 a 20 años de edad, residentes en viviendas públicas. Se realizaron encuestas a los participantes tanto antes como después de la implementación del programa, y éstas mostraron que, antes del programa, un 92% de los encuestados informó que esperaban encontrar alguna clase de problema, y dos terceras partes consideraron que serían víctimas de actos violentos en los próximos tres meses. Después de la implementación del programa, el 20% de los encuestados expresó que esperaban encontrar alguna clase de problema, y sólo un 5% consideró que serían víctimas de actos criminales.

La evaluación del programa de basquetbol nocturno indicó una reducción de un 78% en la tasa de delincuencia juvenil en personas entre 16 y 20 años de edad, en el predio de ubicación para el desarrollo de vivienda pública. La razón primordial de la disminución, según los encuestados, fue que el programa de basquetbol de medianoche les proporcionó una actividad positiva. En este ejemplo, los participantes en el programa son sus beneficiarios directos, quienes piensan que éste les ayuda a mantenerse fuera de problemas y evitar ser víctimas de crímenes violentos.

También se encuestó a residentes de la comunidad, quienes informaron sentir que tanto su comunidad como sus hijos estaban más seguros debido al programa de basquetbol. En este caso, los residentes de la comunidad son beneficiarios indirectos y por lo menos a corto plazo (dependiendo de cuánto duren los beneficios). Aunque no son participantes del programa, se sienten más seguros como resultado de él.

Los hallazgos anteriores podrían utilizarse para demostrar a los residentes y a la comunidad en general que este programa tuvo éxito al prevenir y reducir la violencia, y los administradores del programa podrían también presentar los hallazgos al Concejo

municipal para justificar una solicitud de continuación de la financiación. Estos administradores se constituyen en beneficiarios indirectos si la continuación de la financiación les permite mantener su trabajo un tiempo mayor.

A largo plazo, también la sociedad se beneficia si los jóvenes se mantienen fuera de la cárcel, pues no incurre en los costos de encarcelamiento y de pérdida de productividad. En lugar de ir a la cárcel, los jóvenes quizá se conviertan en ciudadanos que pueden ser empleados, productivos y contribuyentes.

Una evaluación puede, además, suministrar información relacionada con el proceso de ejecución de un programa. Otras agencias públicas de vivienda (beneficiarios no previstos) podrán beneficiarse de las enseñanzas aprendidas durante la fase de ejecución del programa y su posterior evaluación.

Otro ejemplo es el de una intervención para construir y mantener una planta de tratamiento de aguas que produce agua potable para los residentes. El beneficio inmediato es el agua potable limpia para los residentes. Un beneficio directo, a más largo plazo para los miembros de la comunidad y la comunidad en general, sería la reducción en la incidencia de enfermedades transmitidas por el agua. Y beneficios a mediano plazo serían los de atraer industrias a la comunidad por contar ésta con agua potable.

Qué se debe evaluar

En las evaluaciones se consideran diferentes aspectos para el desarrollo (cuadro 1.1). Algunos de los aspectos que pueden evaluarse son los siguientes:

- *Proyectos:* una intervención individual en una ubicación o un proyecto individual ejecutado en varias ubicaciones.
- *Programas:* una intervención que incluye varias actividades o proyectos para contribuir a una meta común.
- *Políticas:* normas, lineamientos o reglas establecidas por una organización para regular decisiones concernientes al desarrollo.
- *Organizaciones:* varios programas de intervención que una organización ejecuta.
- *Sectores:* intervenciones en un campo específico de política, como educación, silvicultura, agricultura o salud.
- *Temas:* asuntos particulares, a menudo de actividades multisectoriales, como equidad de géneros o bienes públicos globales.
- *Ayuda al país:* progreso del país en relación con un plan, efecto general de la ayuda y enseñanzas aprendidas.

Usos de la evaluación

Los resultados de la evaluación tienen diferentes usos. Las evaluaciones proporcionan a clientes, agencias oficiales, organizaciones no gubernamentales (ONG), el público y muchos otros grupos, retroalimentación relacionada con las políticas, programas y

Cuadro 1.1 *Ejemplos de evaluaciones de políticas, programas y proyectos*

	Aplicación	
Tipo de evaluación	*Privatización de sistemas de acueducto*	*Reubicación*
Evaluación de política	Comparar enfoques de modelos para la privatización del suministro público de agua	Comparar estrategias utilizadas para reubicar a los aldeanos en nuevas zonas
Evaluación de programa	Evaluar la gestión fiscal de los sistemas del gobierno	Evaluar el grado al cual los agricultores de la aldea mantienen su medio de vida anterior
Evaluación de proyecto	Comparar el mejoramiento en las tarifas del servicio de agua en dos provincias	Evaluar las prácticas agrícolas de los agricultores reubicados en una provincia

Fuente: los autores.

proyectos. Los resultados suministran información sobre el uso de los fondos públicos, y pueden informar a los gerentes y autores de políticas acerca del correcto –o incorrecto– funcionamiento, en cuanto al cumplimiento de los objetivos originales o revisados.

Las evaluaciones contribuyen a una mayor responsabilidad en la ejecución de proyectos, programas y políticas y en la forma como utilizan los fondos públicos. Con ellas pueden identificarse los proyectos, programas y políticas que justifican replicación, aumento gradual, mejoras o su posible terminación.

Weiss (2004) subrayó la importancia de identificar el uso previsto para una evaluación desde la etapa inicial de planeación. Anota que "Si no es posible identificar y articular usuarios y usos previstos para la evaluación, no debe realizarse ésta", y también que "La evaluación que no se utiliza constituye un desperdicio de valiosos recursos humanos y financieros" (p. 1).

El proceso de evaluación debe diseñarse y realizarse desde el inicio hasta el final, de acuerdo con las necesidades de los usuarios primarios previstos, quienes tendrán la responsabilidad de implementar el cambio, basados en su participación en el proceso o en los resultados de la evaluación.

Las evaluaciones se pueden utilizar para muchos fines (recuadro 1.1); por medio de éstas se puede:

- Contribuir a analizar por qué se obtuvieron, o no, los resultados previstos.
- Explorar por qué hubo resultados o consecuencias imprevistas,
- Evaluar cómo y por qué los resultados se vieron afectados por actividades específicas.
- Ilustrar sobre los procesos de implementación, las fallas y los éxitos que ocurren en cada nivel.

Recuadro 1.1 *Usos de la evaluación*

La evaluación puede utilizarse para:

- Contribuir a la toma de decisiones sobre asignación de recursos.
- Ayudar a reconsiderar las causas del problema.
- Identificar el surgimiento de problemas.
- Apoyar la toma de decisiones en cuanto a alternativas competidoras o mejores.
- Apoyar la reforma del sector público y la innovación.
- Generar consenso sobre las causas de un problema y cómo responder ante ellas.

Fuente: Kusek y Rist, 2004.

- Aportar enseñanzas, destacar los campos de mayor logro y potencial, y ofrecer recomendaciones específicas para el mejoramiento y las reformas.

En resumen, las evaluaciones son útiles al enfocarse en:

- La estrategia política general y los temas de diseño ("¿Estamos haciendo lo correcto?")
- Temas operativos y de implementación ("¿Estamos haciendo las cosas bien?")
- Si existen mejores formas de abordar el problema ("¿Qué estamos aprendiendo?")

Relación entre seguimiento y evaluación

Seguimiento: Recopilación de datos con los cuales los administradores pueden evaluar el grado de cumplimiento de los objetivos.

Con fines de coherencia, se utiliza la definición de *seguimiento* del *Glossary of Key Terms in Evaluation,* de la OCDE/CAD:

> El seguimiento es una función continua en la que se utiliza la recopilación sistemática de datos sobre indicadores especificados con el fin de proporcionar a la administración y las principales partes interesadas de una intervención en curso para el desarrollo, indicaciones del grado de avance y cumplimiento de objetivos y progreso en el uso de los fondos asignados (OCDE, 2002, pp. 27-28).

El seguimiento constituye una actividad de rutina, continua e interna. Se utiliza para recopilar información sobre las actividades, productos y resultados de un programa, a fin de monitorear su funcionamiento.

Como ejemplo de un sistema de seguimiento se presenta el uso, del Ministerio de Salud de Malawi, de 26 indicadores para hacer seguimiento a la calidad de los servicios de salud que se prestan en el Hospital Central. Los indicadores incluyen el número de pacientes atendidos por especialistas en el término de cuatro semanas desde su remisión, el número de muertes de pacientes antes de recibir tratamiento, el número de muertes

directas por obstetricia en el hospital y el número de días de espera de tratamiento de los pacientes (Gobierno de Malawi, 2007).

El suministro regular de datos sobre los indicadores le proporciona al ministro de Salud una línea de tendencia, de modo que pueden investigarse las oscilaciones bruscas. Por ejemplo, un marcado incremento en el número de muertes de pacientes antes de recibir tratamiento puede reflejar una alta tasa de infección en el hospital que es necesario reducir de inmediato. Un marcado descenso en las tasas de infección puede sugerir que el uso de un nuevo desinfectante es eficaz, lo que a su vez sugiere que debe promoverse su uso.

Las evaluaciones se realizan en general para responder a las preguntas "por qué" subyacentes al seguimiento de los datos; preguntas como por qué las operaciones de cesárea aumentan en cinco hospitales y por qué tres de cada 50 sitios presentan tasas de supervivencia particularmente altas para nacimientos prematuros. También se requieren las evaluaciones para atribuir los resultados a una intervención específica y no a otras causas posibles.

Tanto el seguimiento como la evaluación miden y valoran el funcionamiento, pero lo hacen en formas y tiempos distintos (cuadro 1.2).

- El seguimiento se efectúa durante la ejecución de todo el programa o el proyecto.
- La evaluación es la valoración periódica del funcionamiento del programa o el proyecto. Busca responder a la pregunta "¿por qué?"

El seguimiento es una actividad interna que realiza el personal del proyecto, y generalmente es responsabilidad de la gerencia del proyecto verificar que se realice y utilicen sus resultados. En contraste, la evaluación puede realizarse interna o externamente. Es responsabilidad del evaluador y del personal del programa (Insideout, 2005).

Papeles y actividades de los evaluadores

A medida que el concepto y los propósitos de la evaluación evolucionan a través del tiempo, también lo hacen los papeles y actividades de los evaluadores. Éstos representan muchos papeles y se ocupan en numerosas actividades. Su papel depende de la natura-

Cuadro 1.2 *Comparación de seguimiento y evaluación*

Seguimiento	*Evaluación*
En curso, continuo	Periódica y con límite de tiempo
Actividad interna	Interna, externa o participativa
Responsabilidad de la gerencia	Responsabilidad del evaluador y del personal y la gerencia del programa
Retroalimentación continua para mejorar el funcionamiento del programa	Retroalimentación periódica

Fuente: Insideout, 2005.

leza y el propósito de la evaluación. Como mencionan Fitzpatrick, Sanders y Worthen (2004, p. 28), "los evaluadores tienen muchos papeles, entre ellos los de experto científico, facilitador, planificador, colaborador, asistente de quienes toman las decisiones y amigo crítico". Además, actúan como jueces, personas de confianza, maestros y agentes del cambio social.

¿Quién realiza la evaluación?

Los evaluadores pueden formar parte de evaluaciones internas, externas o participativas. El glosario de la OCDE/CAD define la *evaluación interna* como:

> Evaluación de una intervención de desarrollo realizada por una unidad o individuos que reportan a la gerencia de la organización donante, asociada o ejecutora (2002, p. 26).

Define *evaluación externa* como:

> Evaluación de una intervención de desarrollo realizada por entidades y/o individuos de fuera de la organización donante, asociada o ejecutora (2002, p. 23).

Evaluación interna: Realizada por personal interno de la organización.

Evaluación externa: Realizada por personas de fuera de la organización.

Las evaluaciones internas y externas tienen sus ventajas y desventajas. Por lo general, los evaluadores internos conocen más el programa, el proyecto o la política que los externos. La persona que desarrolla y administra la intervención puede ser también la responsable de la evaluación. Estas personas suelen conocer más la historia, la organización, la cultura, los participantes, los problemas y los éxitos. Debido a este conocimiento, los evaluadores internos pueden hacer preguntas más importantes y pertinentes; saben a dónde acudir entre bastidores en la organización para averiguar lo que realmente está sucediendo.

Sin embargo, esta ventaja puede también ser una desventaja. Los evaluadores internos pueden estar tan cerca del programa, el proyecto o la política, que no lo ven con claridad, y es posible que no reconozcan soluciones o cambios que otros sí pueden apreciar. Los evaluadores internos tienen también la desventaja de estar más sujetos a presiones o influencias de quienes toman las decisiones sobre el programa, que son también quienes toman las decisiones sobre el personal. Ven la organización total sólo desde su posición dentro de ella. Además, las partes interesadas externas perciben sus hallazgos como menos verosímiles que los de los evaluadores externos.

Los evaluadores externos casi siempre tienen más credibilidad y aportan una percepción de objetividad a una evaluación. Además, la mayoría de los evaluadores externos cuentan con destrezas más especializadas requeridas por las evaluaciones eficaces y también son independientes de la administración y de las decisiones financieras sobre el programa (Fitzpatrick, Sanders y Worthen, 2004).

No obstante, una evaluación externa no constituye garantía de resultados independientes y verosímiles, en particular si los consultores han tenido lazos anteriores con el programa. Los consultores externos son más complacientes con la administración por la esperanza de obtener más trabajo en el futuro.

La evaluación participativa se considera cada vez más como un tercer método de evaluación. Los evaluadores participativos trabajan en conjunto con los representantes de las agencias y las partes interesadas para diseñar, realizar e interpretar una evaluación (OCDE, 2002). La evaluación participativa difiere de la evaluación interna y externa en algunas formas fundamentales.

> La evaluación participativa representa un paso adicional y más radical que se aleja del modelo de la evaluación independiente… es una forma de evaluación en la que se resta énfasis y se redefine la distinción entre expertos y legos, investigadores e investigados… Los evaluadores [actúan] principalmente [como] facilitadores e instructores, ayudando a otros a hacer la valoración (Molund y Schill, 2004, p. 19).

Debe anotarse la distinción entre evaluación participativa y métodos participativos, que pueden utilizarse tanto en las evaluaciones internas como en las externas.

Actividades del evaluador

Los evaluadores realizan las actividades correspondientes a sus distintos papeles. Los evaluadores internos trabajan en el diseño del programa o el proyecto, su ejecución y en estrategias de alcance. Los evaluadores externos usualmente limitan su participación en la administración del programa. Por lo general, todos los evaluadores:

- Consultan a todas las partes interesadas importantes.
- Administran presupuestos de evaluación.
- Planifican la evaluación.
- Realizan la evaluación o contratan personal para hacerla.
- Identifican estándares para la eficacia (basados en documentos de autorización u otras fuentes).
- Recopilan, analizan, interpretan e informan sobre datos y hallazgos.

Para cumplir sus metas los evaluadores requieren diversas destrezas. Como parte del proceso de evaluación, contribuyen a generar conocimiento y difundir las enseñanzas adquiridas.

Orígenes e historia de la disciplina de evaluación

La moderna disciplina de la evaluación tuvo su origen en la investigación en las ciencias sociales, basada en el método científico, pero la evaluación tiene antiguas tradiciones. En realidad, la evidencia arqueológica muestra que los antiguos egipcios efectuaban

seguimiento con regularidad a la producción de granos de su país y la producción de ganado hace más de 5.000 años. En la China y Grecia antiguas también se llevaban a cabo evaluaciones:

> En el sector público, la evaluación formal era evidente ya desde el año 2000 a.C., cuando los funcionarios chinos efectuaban exámenes del servicio civil para medir la competencia de los aspirantes a posiciones oficiales. En educación, Sócrates utilizó evaluaciones verbales como parte del proceso de aprendizaje (Fitzpatrick, Sanders y Worthen, p. 31).

Algunos expertos ubican el surgimiento de los métodos modernos de evaluación en la aparición de las ciencias naturales y el énfasis en la observación de los fenómenos (el método empírico) en el siglo XVII. En Suecia, en ese tiempo aparecieron comisiones con fines específicos de política que efectuaban alguna clase de evaluación. De hecho, el sistema de comisiones subsiste todavía en Suecia y actualmente hay varios centenares de comisiones.

En el siglo XIX, la evaluación de programas educativos y sociales empezó a cimentarse en varios países anglosajones. La evaluación de programas se realizaba en Gran Bretaña por medio de comisiones nombradas por el gobierno, a las que se les solicitaba investigar y evaluar la insatisfacción con los programas educativos y sociales. Los actuales órganos de inspección externa para los colegios surgieron de estas primeras comisiones.

En los Estados Unidos se realizaron esfuerzos pioneros durante el siglo XIX con el fin de analizar la calidad del sistema escolar por medio de pruebas de logros. Estos esfuerzos continúan hasta el presente y los puntajes de las pruebas de los estudiantes siguen siendo una medida básica para establecer la calidad de la educación en los colegios. En los Estados Unidos también en este período se inició la acreditación para las instituciones de secundaria y las universidades.

Historia de la evaluación en el siglo XX

La evaluación formal y la acreditación en los colegios de medicina de los Estados Unidos y Canadá se establecieron a comienzos del siglo XX. Durante este período otros campos de la investigación, la medición y la evaluación fueron los de la salud, vivienda, productividad laboral, liderazgo democrático y autoritario y las pruebas educativas estandarizadas. La mayoría sólo fueron esfuerzos a pequeña escala por parte de agencias oficiales y servicios sociales.

Rossi y Freeman (1993) señalan los orígenes del compromiso con la evaluación sistemática de programas en los Estados Unidos en los esfuerzos de finales del siglo XIX y principios del XX para mejorar la alfabetización, proporcionar capacitación ocupacional y reducir las muertes por enfermedades infecciosas. En el campo del desarrollo, el "intento de introducir el agua hervida como práctica de salud pública en las aldeas en el Medio Oriente constituye un hito en los estudios de la literatura sociológica empírica anterior a la Segunda Guerra Mundial" (Rossi y Freeman, 1993, p. 10).

La investigación social aplicada aumentó rápidamente en los Estados Unidos después de la institución del *New Deal* del presidente Franklin D. Roosevelt. El gobierno federal creció con rapidez y se fundaron nuevas agencias para manejar e implementar programas nacionales. Estos programas incluyeron subsidios para los agricultores, obras públicas y planes de creación de empleo, electrificación rural y seguridad social. Dado que los programas en gran escala constituían una novedad y su naturaleza era experimental, aumentó la necesidad de evaluar su eficacia para arrancar la economía, crear empleos y redes de seguridad social.

La necesidad de la evaluación aumentó durante la Segunda Guerra Mundial y después de ella, al diseñarse e iniciarse un mayor número de programas de gran escala para el ejército, la vivienda urbana, la capacitación laboral y ocupacional, y la salud. Durante este tiempo también se realizaron los mayores compromisos con programas internacionales que incluyeron planificación familiar, salud y nutrición y desarrollo comunitario rural. Los gastos fueron grandes y, en consecuencia, estuvieron acompañados de demandas por conocer los resultados.

En los años cincuenta y sesenta se utilizó la evaluación de forma más rutinaria en los Estados Unidos y Europa, con el fin de evaluar programas relacionados con la educación, la salud física y mental, los servicios humanos, la prevención de la delincuencia y la rehabilitación de los delincuentes. El programa de "Guerra contra la pobreza", del presidente Lyndon B. Johnson durante los años sesenta, estimuló un mayor interés en la evaluación. Igualmente se expandió el trabajo en los países en desarrollo alrededor del mundo, realizándose alguna evaluación de programas agrícolas, de desarrollo comunitario, planificación familiar, servicios de salud y nutrición. En su mayor parte, estas evaluaciones contaban con las herramientas tradicionales de las ciencias sociales, como las encuestas y el análisis estadístico.

En 1949, la primera Comisión Hoover recomendó que la información presupuestaria para el gobierno nacional de los Estados Unidos se estructurase en términos de actividades, en lugar de renglones. Recomendó además que se proporcionasen medidas de desempeño en los informes (Burkhead, 1956; Mikesell, 1995). Este tipo de reforma presupuestaria se conoció como *presupuesto de desempeño* (Tyler y Willand, 1997).

En 1962, el Departamento de Defensa de EUA, con el secretario Robert McNamara, desarrolló el Sistema de Planeación, Programación y Presupuesto (PPBS), cuyo propósito fue incrementar la eficiencia y mejorar las operaciones del gobierno. Esto implicó:

- Establecer objetivos de planeación a argo plazo.
- Analizar los costos y beneficios de programas alternativos que cumpliesen estos objetivos.
- Trasladar los programas a presupuestos y propuestas legislativas, y proyecciones a largo plazo.

El PPBS representó un cambio en el tradicional proceso de presupuesto al hacer hincapié en los objetivos y vincular la planeación y el presupuesto (Office of the Secretary of Defense, 2007). Los primeros esfuerzos del PPBS llevarían eventualmente al movimiento del "seguimiento para resultados".

A finales de los años sesenta, muchos países de Europa occidental empezaron a desarrollar evaluación de programas. Así, en la República Federal de Alemania el Bundestag comenzó a exigirle al gobierno federal informes sobre la ejecución y el impacto de los distintos programas socioeconómicos y tributarios, algunos de los cuales eran los referentes a la Ley sobre el Mercado Laboral y el Empleo (1969), la Ley General de Garantías Educativas (1969), el programa conjunto de los estados y la federación para mejorar la estructura económica regional (1970), el programa de inversiones hospitalarias (1971) y varios informes sobre subsidios e impuestos (Derlien, 1999). Durante este período, el gobierno canadiense empezó también a realizar evaluaciones de los programas del gobierno y su funcionamiento, y se alentó a los departamentos del gobierno a establecer unidades de planeación y evaluación.

Los primeros esfuerzos no produjeron resultados significativos, y en Canadá, la República Federal de Alemania y Suecia, "a pesar de la institucionalización de la evaluación de programas en varios campos de la política, la fragmentación continuó siendo característica de sus sistemas y el número de estudios realizados fue relativamente bajo" (Derlien, 1999, p. 146).

La Ley sobre Educación Primaria y Secundaria (ESEA) de 1965 marcó un hito para la evaluación en los Estados Unidos, y ordenaba por parte del gobierno que se evaluara el desempeño de los estudiantes y los estándares de calidad para los maestros. Además proporcionó recursos (el primer renglón especial en el presupuesto del gobierno de EUA para evaluación) para iniciar estas actividades, institucionalizando con esto la evaluación. Al fluir el dinero federal hacia las evaluaciones a finales de los años sesenta y principios de los setenta, empezaron a aparecer numerosos artículos y libros en los Estados Unidos y algunos países de la OCDE y se desarrollaron programas universitarios de posgrado sobre evaluación con el fin de ofrecer capacitación para grupos de base de evaluadores, y satisfacer la creciente demanda de rendición de cuentas y eficacia en los programas socioeconómicos financiados por el gobierno, como las becas educativas para estudiantes de primaria y secundaria, y programas de la "gran sociedad", entre ellos los de reducción de la pobreza, preescolares del programa Head Start (ventaja inicial), actividades de derechos civiles y la creación de un órgano de empleo.

En Canadá, la República Federal de Alemania y Suecia se inició la evaluación de programas en la década de 1960 con los nuevos programas de educación, salud y bienestar social financiados por el gobierno.

> En este contexto surgieron los sistemas formales de planeación, que se limitaron a la planeación financiera a mediano plazo (en la República Federal de Alemania), o a integrar el presupuesto con la programación (en Suecia y Canadá). De todos modos, se consideró que la evaluación formaba lógicamente parte de estos sistemas de planeación o que era necesaria

> para satisfacer las necesidades de información de los programas de intervención… Entonces, las evaluaciones fueron utilizadas principalmente por los administradores de los programas en la ejecución de los programas nuevos y existentes (Derlien, 1999, pp. 153-54).

A partir de mediados de los años setenta y hasta mediados de los ochenta, la evaluación constituyó una profesión con todas las de la ley en muchos países de la OCDE. Se crearon asociaciones de evaluación profesional, se introdujeron nuevos programas de capacitación para evaluadores y también revistas especializadas, y la evaluación empezó a extenderse más allá de la esfera de los programas financiados por el gobierno, a corporaciones, fundaciones e instituciones religiosas. Por ejemplo, en Francia se desarrolló sistemáticamente la evaluación de políticas públicas y muchas universidades –incluidas las *Grandes Écoles*– empezaron a ofrecer cursos e información sobre la evaluación como parte de sus planes de estudio.

Muchos países de la OCDE han instituido programas de capacitación en evaluación para los funcionarios públicos, ya sea dentro del gobierno o por medio de contratistas externos. Se exploraron nuevas metodologías y modelos, con mayor énfasis en las necesidades de información de los consumidores, el examen de resultados no previstos y el desarrollo de valores y estándares. La literatura sobre la evaluación creció también tanto en cantidad como en calidad (Fontaine y Monnier, 2002). A partir de 1985, los computadores y la tecnología han contribuido a aumentar en gran medida la capacidad de los evaluadores para recopilar, analizar e informar los resultados de las evaluaciones y compartirlos con otros.

Evaluación para el desarrollo y auditoría

La evaluación para el desarrollo evolucionó de forma que se apartó de la tradición de la auditoría y las ciencias sociales, y existen importantes similitudes, diferencias y vínculos entre las dos tradiciones.

La tradición de la auditoría

Los orígenes de la auditoría se remontan a la Inglaterra del siglo XIX, cuando el creciente desarrollo comercial e industrial generó

> la necesidad de registros financieros de precisión y confiabilidad comprobables… El trabajo de los auditores le aportó credibilidad a la creciente infraestructura capitalista en Occidente. La opinión de los auditores era respetada por su destreza técnica y porque los auditores eran externos (Brooks, 1996, p. 16).

La tradición en la auditoría tiene una orientación investigadora, de administración financiera y contable. Busca establecer si en un programa se ejecutó lo debido y si el dinero se utilizó conforme a las reglas, regulaciones y requerimientos del programa. Utiliza conceptos como el de control interno, la buena gestión y la gobernanza, y la

verificación. Su énfasis está en la rendición de cuentas y el cumplimiento. El glosario de la OCDE/CAD (OCDE, 2002) define varios tipos de auditoría:

- *Auditoría estándar:* Actividad de seguridad independiente y objetiva, diseñada para agregar valor y mejorar las operaciones de una organización (contribuye a lograr los objetivos de una organización, aportando un enfoque sistemático y disciplinado, y a mejorar la eficacia de los procesos de gestión del riesgo, el control y gobernanza).
- *Auditoría financiera:* Auditoría centrada en el cumplimiento de los estatutos y regulaciones aplicables.
- *Auditoría del desempeño:* Auditoría referente a la pertinencia, la economía, la eficiencia y la eficacia.

Auditoría estándar: Actividad de seguridad independiente y objetiva diseñada para mejorar las operaciones de una organización.

Auditoría financiera: Centrada en el cumplimiento de los estatutos y regulaciones aplicables.

Auditoría del desempeño: Evalúa la pertinencia, la economía, la eficiencia y la eficacia.

La función de la auditoría puede ser interna o externa (recuadro 1.2). Como función interna, la auditoría contribuye a lograr los objetivos de una organización, aportando un enfoque sistemático y disciplinado para evaluar y mejorar la eficacia de los procesos de gestión del riesgo, el control y la gobernanza. Los gobiernos utilizan auditores externos cuando se requiere la independencia de la administración de un programa.

La fuerza de la tradición de la auditoría en los países en desarrollo condujo también a una fuerte tradición de auditoría sobre el cumplimiento en la evaluación. Así, por ejemplo, el Departamento Nacional de Auditoría de Malasia (NAD, por sus siglas en

Recuadro 1.2 *El Institute of Internal Auditors y la International Organization of Supreme Audit Institutions*

El Institute of Internal Auditors fue fundado en 1941, y constituye la asociación internacional de primer orden dedicada a la promoción y el desarrollo de la auditoría interna. Existen institutos nacionales en todo el mundo y sus integrantes provienen de 165 países.

En 1947, el Institute of Internal Auditors publicó *The Statement of Responsibilities of Internal Auditing,* que se convirtió en la base para desarrollar los estándares de auditoría interna. En 1978, se aprobaron los *Standards for the Professional Practice of Internal Auditing.*

Desde Afganistán hasta Zimbabue, casi todos los países desarrollados y en desarrollo pertenecen a la Organización Internacional de Entidades Fiscalizadoras Superiores (INTOSAI), que opera como una organización de cobertura para la comunidad de auditoría oficial externa. Fundada en 1953, la INTOSAI promueve el desarrollo y la transferencia del conocimiento para mejorar la auditoría en el gobierno en todo el mundo. Su *Code of Ethics and Auditing Standards* y su *Implementation Guidelines for Performance Auditing* son de uso general. En 2008, había 188 países miembros de Indosai.

Fuente: http://www.theiia.org; http://www.intosai.org.

inglés) ha venido asegurando la rendición de cuentas pública durante 100 años en ese país. El NAD realiza auditorías para:

- Asegurar el cumplimiento de las leyes y regulaciones.
- Exponer factores injustificados que ocasionan ineficiencia, ineficacia o procedimientos no económicos.
- Determinar si los estados financieros son correctos y razonables, y si los registros se prepararon de acuerdo con los principios generalmente aceptados de contabilidad y auditoría.

La **auditoría interna** comprende una gran variedad de "actividades y operaciones financieras que incluye sistemas, producción, ingeniería, mercadeo y recursos humanos" (http://www.theiia.org), y su fortaleza reside en el hecho de que se ofrece acreditación profesional, lo que no sucede aún con la evaluación (en el capítulo 15 se tratan los pros y contras de la acreditación para evaluadores).

Auditoría interna: La que se efectúa dentro de una organización para evaluar actividades y operaciones financieras.

La evaluación para el desarrollo toma de la profesión de la auditoría un fuerte enfoque en el cumplimiento de los requisitos legales y procedimentales, lo cual puede observarse en los marcos de evaluación de proyectos basada en objetivos de las entidades donantes bilaterales y los bancos de desarrollo. Por ejemplo, los "Good Practice Standards for Evaluation of MDB-Supported Public Sector Operations", desarrollados por el Grupo de Cooperación para la Evaluación de los Banco(s) Multilateral(es) de Desarrollo (BMD), incluyen el cumplimiento de los objetivos, con base en que "la evaluación contra los objetivos mejora la responsabilidad" (Evaluation Cooperation Group of the Multilateral Development Banks, sin fecha, p. 9).

El continuo entre la auditoría y la evaluación

Puede considerarse que la auditoría y la evaluación forman parte de un continuo que suministra información relacionada, aunque diferente, sobre el cumplimiento, la rendición de cuentas, el impacto y los resultados de los proyectos y programas. Existe alguna "superposición en campos como la eficiencia en las operaciones y la efectividad en los costos… tratando la evaluación sobre el análisis de políticas y los productos y la auditoría sobre los controles financieros internos y los sistemas de gestión" (Treasury Board of Canada Secretariat, 1993, p. 3). Tanto la auditoría como la evaluación están encaminadas a servir de ayuda a quienes toman las decisiones "suministrándoles información sistemática y verosímil, útil para la creación, la gestión, la supervisión, el cambio y, ocasionalmente, la revocación de los programas" (Wisler, 1996, p. 1).

Se ha escrito mucho acerca de las diferencias entre la auditoría y la evaluación y la superposición entre ambas. Las diferencias se encuentran en sus orígenes; la auditoría se deriva en gran medida de la contabilidad financiera y la evaluación de las ciencias sociales. La auditoría suele centrarse en el cumplimiento de los requisitos, en tanto que la evaluación lo suele hacer en atribuir los cambios observados a una política, un programa

o un proyecto. Los auditores tienden a buscar respuestas a cuestiones normativas (lo que es frente a lo que debería ser), mientras los evaluadores tienden a buscar respuestas a cuestiones descriptivas y de causa y efecto (Wisler, 1996).

La tradición de las ciencias sociales

Al pasar los gobiernos y las organizaciones de enfatizar la verificación y el cumplimiento a hacer hincapié en el impacto, se incorporaron las técnicas de las ciencias sociales a la evaluación. La evaluación para el desarrollo aprovechó el método científico y la investigación social.

El método científico constituye un enfoque sistemático de adquisición de información que busca de manera objetiva separar la verdad de la creencia. Según este enfoque se recopilan los datos por medio de la observación y la experimentación, y se basan en la formulación y el sometimiento a prueba de hipótesis. Los investigadores que utilizan el método científico:

- Identifican un problema, lo investigan y analizan explicaciones anteriores.
- Desarrollan una hipótesis en cuanto a una relación de causa y efecto, y la expresan en términos cuantificables.
- Someten a prueba la hipótesis, realizando experimentos y recopilando datos.
- Analizan los datos y sacan una conclusión.
- Documentan y difunden sus hallazgos.
- Utilizan los resultados para perfeccionar la hipótesis.

La evaluación aprovechó otros métodos de las ciencias sociales, como la sociología, la antropología, la ciencia política y la economía.

> La aplicación de los métodos de investigación de las ciencias sociales a la evaluación coincide con el crecimiento y el perfeccionamiento de los métodos mismos, y también con los cambios ideológicos, políticos y demográficos ocurridos durante este siglo. Fueron de la mayor importancia el surgimiento y la posterior posición destacada de las ciencias sociales en las universidades y el mayor apoyo que se le dio a la investigación social. Los departamentos de ciencias sociales en las universidades se convirtieron en centros de trabajo pionero en la evaluación de programas y continúan ocupando un lugar influyente en este campo (Rossi y Freeman, 1993, p. 9).

La evaluación también hace muy buen uso de la investigación de las ciencias sociales en campos como la construcción de teorías, el diseño, el enfoque, la metodología de recopilación de datos, su análisis e interpretación, la estadística, las encuestas y el muestreo.

El contexto de la evaluación para el desarrollo

La evaluación para el desarrollo surgió como una subdisciplina de la evaluación, y empezó principalmente a partir de los esfuerzos de reconstrucción y desarrollo que se efectuaron al terminar la Segunda Guerra Mundial. En 1944, se creó el Banco Mundial y su primer

grupo de evaluación independiente se estableció en 1972. En 1991, se fundó el Banco Europeo para la Reconstrucción y el Desarrollo (BERD) y también se fundaron otros bancos de desarrollo multilateral en la década de 1990. Además, se establecieron organizaciones bilaterales, como el Department for International Development (DfID), del Reino Unido, y la Canadian International Development Agency (CIDA). Los orígenes de la evaluación para el desarrollo estuvieron en el establecimiento de estas organizaciones, al ser los donantes los responsables de la financiación de los proyectos y de sus resultados. A su vez, los ministerios de los países en desarrollo debían cumplir los requerimientos de informar sobre los hallazgos de los proyectos, utilizando sistemas de evaluación de proyectos desarrollados por las entidades donantes para el aprendizaje y la contabilización.

A medida que ha ido cambiando la noción de desarrollo en las últimas décadas, también lo ha hecho la evaluación para el desarrollo. Por ejemplo, desde sus inicios, el Banco Mundial ha variado su énfasis, con implicaciones en la complejidad de la evaluación para el desarrollo (cuadro 1.3).

La OCDE ha desempeñado un papel importante en el avance de la evaluación para el desarrollo. Fundada en 1961, su misión ha sido la de "ayudar a los gobiernos a obtener el desarrollo económico sostenible y el empleo, y aumentar los niveles de vida en los países integrantes, manteniendo la estabilidad fiscal, contribuyendo así al desarrollo de la economía mundial" (http://www.ocde.org). Los integrantes de la OCDE se reúnen en comités especializados, incluyendo el Comité de Asistencia para el Desarrollo (CAD), que ha constituido por mucho tiempo un grupo de trabajo sobre la evaluación para el desarrollo, actualmente, la CAD Network on Evaluation, cuyo propósito es aumentar la eficacia de los programas de desarrollo internacional con el apoyo de una evaluación sólida, informada e independiente. Reúne a 30 agencias de desarrollo bilateral y multilateral.

Si bien se ha adoptado ampliamente la definición del CAD de evaluación para el desarrollo, ésta difiere algo de la definición genérica de evaluación que se presentó al principio de este capítulo. De acuerdo con el CAD, una evaluación para el desarrollo es:

> La valoración sistemática y objetiva del diseño, ejecución y resultados de un proyecto, un programa o una política en curso o terminada. Su propósito es determinar la aplicabilidad y el cumplimiento de los objetivos, la eficiencia y la eficacia para el desarrollo, su impacto y su sostenibilidad. Una evaluación debe suministrar información verosímil y útil, habilitando la incorporación de las enseñanzas adquiridas en el proceso de toma de decisiones de los receptores y de los donantes (OCDE, 1991b, p. 4).

La comunidad de evaluación utiliza una amplia variedad de metodologías y prácticas y en general se acepta, ya que una combinación de teorías, estrategias de análisis y metodologías con mucha frecuencia funciona mejor en la evaluación para el desarrollo, especialmente dada la escala creciente y la complejidad de los proyectos, programas y políticas para el desarrollo. La combinación de enfoques contribuye a fortalecer la evaluación.

Esta combinación de métodos, llamada *triangulación metodológica,* se refiere a:

Cuadro 1.3 *El cambio en el enfoque del Banco Mundial sobre el desarrollo, 1950-2000*

Década	*Foco*	*Enfoque*	*Disciplina*
1950	Reedificación, reconstrucción, asistencia técnica e ingeniería	Asistencia técnica	Ingeniería
1960	Crecimiento económico, financiación y creación de proyectos, con la esperanza de que un crecimiento económico más fuerte eliminara la pobreza de muchas personas	Proyectos	Finanzas
1970	Sectores sociales o necesidades básicas (educación, salud y bienestar social); planeación a largo plazo e inversiones en el sector social	Inversión sectorial	Planeación
1980	Políticas de ajuste estructural y préstamos; vinculación de créditos de ajuste a condiciones específicas utilizadas en apoyo de reformas políticas importantes y para ayudar a los países a hacer frente a las crisis financieras y de endeudamiento	Préstamos de ajuste	Economía neoclásica
1990	Proyectos integrales con base en los países en contraposición a proyectos individuales; mayor énfasis en la formación de capacidades y en las instituciones en los países en desarrollo.	Asistencia a países	Multidisciplinaria
2000	Reducción de la pobreza, asociaciones, participación, enfoques sectoriales y orientación a resultados	Asociaciones	Administración basada en resultados

Fuente: basado en Picciotto, 2002.

> El uso de varias teorías, fuentes o tipos de información y/o tipos de análisis para verificar y sustanciar una valoración. Al combinar múltiples fuentes de datos, métodos, análisis o teorías, los evaluadores buscan superar el sesgo proveniente de informantes individuales, métodos individuales, observadores individuales, o estudios de una teoría individual (OCDE, 2002, p. 37).

Según anotan Chelimsky y Shadish (1997, p. 11):

> La evaluación se hace cada vez más diversa metodológicamente. En la actualidad, se encuentra bien aceptado que todos los métodos de las ciencias sociales hacen parte de la caja de herramientas metodológica del evaluador; esto es, las herramientas de la psicología, la estadística, la educación, la sociología, la ciencia política, la antropología y la economía.

La elección del diseño de evaluación y su metodología (o la combinación de diseños y metodologías) estará determinada por las preguntas que se planteen y la información que se busque obtener.

Crecimiento de asociaciones de evaluación profesional

Las asociaciones de evaluación profesional crean un sistema de apoyo, y permiten el profesionalismo dentro de la comunidad de evaluación. Este apoyo contribuye a desarrollar capacidades en la evaluación para el desarrollo.

En los años ochenta únicamente existían tres sociedades de evaluación regional o nacional, pero desde entonces el crecimiento en nuevas asociaciones de evaluación nacional, regional e internacional ha sido explosivo, y éstas han brotado alrededor del mundo; en la actualidad existen más de 75 asociaciones de evaluación en los países en desarrollo y desarrollados (http://www.ioce.net/www.ioce.net/members/eval_associations.shtml). Gran parte del crecimiento proviene de la fundación de asociaciones de evaluación en países en desarrollo. Por ejemplo, a nivel nacional, han existido asociaciones en Malasia y Sri Lanka desde 1999, y en Uganda desde 2002. A nivel regional, la Australasian Evaluation Society fue fundada en 1991, la European Evaluation Society en 1994 y la African Evaluation Association en 1999.

Una organización internacional importante para la evaluación es la Organización Internacional para la Cooperación en Evaluación (IOCE), una alianza libre de organizaciones de evaluación regional y nacional (asociaciones, sociedades y redes) de todo el mundo. Su propósito es formar liderazgo y capacidad en evaluación en países en desarrollo y fomentar el intercambio mutuamente productivo y benéfico de la teoría y la práctica de la evaluación alrededor del mundo. Para conseguirlo, los profesionales de la evaluación deben adoptar un enfoque más global para contribuir a la identificación y la solución de los problemas mundiales (http://ioce.net/overview/general.shtml).

Otra organización internacional importante para la evaluación es la Asociación Internacional de Desarrollo de Evaluación (IDEAS), creada en 2001 para ayudar a capacitar en evaluación en países en desarrollo. Su misión es "avanzar y extender la práctica de la evaluación para el desarrollo mediante el perfeccionamiento de los métodos, el fortalecimiento de capacidades y la expansión de la propiedad" (http://www.ideas-int.org). IDEAS es la única asociación para profesionales practicantes de evaluación para el desarrollo.

La estrategia de IDEAS es:

- Promover la evaluación para el desarrollo para obtener resultados, transparencia y rendición de cuentas en las políticas y el gasto públicos.
- Dar prioridad al desarrollo de capacidades en evaluación.
- Fomentar los más altos estándares intelectuales y profesionales en la evaluación para el desarrollo.

Principios y estándares para la evaluación para el desarrollo

La comunidad de evaluación requiere principios y estándares que promuevan la rendición de cuentas, faciliten la comparación y mejoren la confiabilidad y calidad de los servicios (Picciotto, 2005). La mayoría de las organizaciones relacionadas con el

desarrollo emplean los principios, estándares y criterios de la OCDE/CAD, y en un documento básico (OCDE, 1991a) se identifican cinco criterios para evaluar la ayuda para el desarrollo; éstos son:

- *Pertinencia:* Grado al cual los objetivos de una intervención de desarrollo son coherentes con los requerimientos de los beneficiarios, las necesidades del país, las prioridades globales y las políticas de las agencias asociadas y de desarrollo.
- *Eficacia:* Medida del grado al cual una actividad de ayuda obtiene sus objetivos.
- *Eficiencia:* Medida –cualitativa y cuantitativa– de los productos en relación con los insumos (este término económico significa que la ayuda utiliza los recursos menos costosos posibles para obtener los resultados deseados. La medición de la eficiencia requiere generalmente que se comparen distintos enfoques alternativos para obtener los mismos productos, a fin de determinar si se empleó el proceso más eficiente).
- *Impacto:* Los cambios positivos y negativos producidos por una intervención de desarrollo, directa o indirectamente, previstos o imprevistos. (La medición del impacto incluye determinar los impactos y efectos principales de una actividad en los indicadores sociales, económicos, ambientales y otros indicadores de desarrollo. El examen debe tratar resultados previstos e imprevistos e incluir el impacto negativo y positivo de factores externos, como cambios en las condiciones comerciales y financieras.)
- *Sostenibilidad:* La flexibilidad con respecto al riesgo de los flujos netos de beneficio a través del tiempo (la noción de sostenibilidad es particularmente pertinente para evaluar [no medir] si existe probabilidad de que los beneficios de una actividad o un programa continúen después de retirarse la financiación de la entidad donante. Los proyectos y programas deben ser sostenibles tanto ambiental como financieramente (OCDE, 1991b).

El CAD desarrolló principios para la evaluación de la ayuda para el desarrollo (OCDE, 1991b), que tratan sobre los temas siguientes:

- Propósito de la evaluación.
- Imparcialidad e independencia.
- Credibilidad.
- Utilidad.
- Participación de donantes y receptores.
- Cooperación de donantes.
- Programación de la evaluación.
- Diseño e implementación de evaluaciones.
- Presentación, difusión y retroalimentación de información.
- Aplicación de estos principios.

En 1998, se efectuó una revisión de los "Principles for Evaluation of Development Assistance" del CAD, en la que se compararon estos principios con los de otras organizaciones, y se buscaron consistencia y posibles campos para la expansión. Las recomendaciones de los miembros para posibles modificaciones a los principios incluyeron las siguientes:

- Modificar la declaración de propósitos.
- Tratar directamente la cuestión de las evaluaciones descentralizadas y las evaluaciones participativas.
- Mayor elaboración sobre los principios y prácticas para la participación de los receptores y la cooperación de las entidades donantes.
- Introducir los desarrollos recientes en la actividad de evaluación, como medidas de desempeño, estado y sistemas de calificación de éxito, y desarrollar una tipología de actividad de evaluación (OCDE, 1998).

Esta revisión sentó las bases para publicaciones posteriores del CAD.

En 1994, la American Evaluation Association (AEA) publicó su *Program Evaluation Standards in the United States,* estándares que fueron aprobados por el American National Standards Institute (ANSI) como los American National Standards for Program Evaluation, fueron actualizados en 1998 y han sido adaptados por otras asociaciones, incluidas las de los países en desarrollo.

En marzo de 2006, la CAD Evaluation Network instituyó los "CAD Evaluation Quality Standards" (OCDE, 2006) (recuadro 1.3). Con estos estándares, utilizados actualmente a manera de ensayo para aplicación en la fase de pruebas, se identifican los pilares requeridos por un proceso de evaluación de calidad y producto:

- Proporcionar estándares para el proceso (conducta) y productos (resultados) de las evaluaciones.
- Facilitar la comparación de evaluaciones entre países (metaevaluación).
- Facilitar asociaciones y colaboración en evaluaciones conjuntas.
- Capacitar mejor a los países miembros para utilizar los hallazgos y los informes de las evaluaciones de otro país (incluyendo buenas prácticas y enseñanzas adquiridas).
- Modernizar los esfuerzos de evaluación.

A solicitud de la OCDE/CAD, la Evaluation Network y otras redes de evaluación, el Independent Evaluation Group del Banco Mundial desarrolló principios y estándares de consenso indicativos para evaluar el Global and Regional Partnership Program (GRPP), que contiene algunas características únicas que hacen compleja la evaluación. Estos principios y estándares indicativos se están sometiendo a prueba con el uso, y serán modificados y aprobados en el término de pocos años. (El enlace al sitio web para estos principios y estándares aparece al final del presente capítulo, y los principios y estándares se tratan con mayor detalle en el capítulo 14.)

001El camino.indd 26 08/11/2010 03:27:41 p.m.

Recuadro 1.3 *Las 10 partes de los estándares de calidad de evaluación del CAD*

Para valorar la calidad de la evaluación, la OCDE fijó 10 criterios que se presentan a continuación:

- Fundamento, propósito y objetivos de una evaluación
- Alcance de la evaluación
- Contexto
- Metodología de la evaluación
- Fuentes de información
- Independencia
- Ética de la evaluación
- Garantía de la calidad
- Aplicabilidad de los resultados de la evaluación
- Integridad

Fuente: OCDE, 2006.

Evaluación independiente: Efectuada por personas sin obligación con quienes diseñaron y ejecutaron la intervención.

Un componente importante de credibilidad de la evaluación para el desarrollo es la independencia, y el glosario OCDE/CAD define la evaluación independiente como "una evaluación realizada por entidades y personas libres del control de los responsables del diseño y ejecución de la intervención de desarrollo" (OCDE, 2002, p. 25). En el glosario se anota lo siguiente:

> La credibilidad de una evaluación depende en parte de la independencia con que se haya efectuado. La independencia implica libertad de influencia política y presión organizativa y se caracteriza por el acceso total a la información y la completa autonomía para llevar a cabo investigaciones e informar sobre los hallazgos.

La independencia no significa aislamiento: la interacción entre los evaluadores, los gerentes de programa, el personal y los beneficiarios pueden mejorar la evaluación y su uso. Una evaluación puede ser llevada a cabo, interna o externamente, por evaluadores que organizativamente están bajo quienes son responsables de tomar las decisiones sobre el diseño y la ejecución de las intervenciones programáticas (es decir, la administración). Dichas evaluaciones no constituyen evaluaciones independientes y sirven para propósitos de aprendizaje, pero no de rendición de cuentas.

Los jefes de evaluación de los bancos de desarrollo multilateral, quienes regularmente se reúnen como miembros del Evaluation Cooperation Group, han identificado cuatro dimensiones o criterios para la independencia de las evaluaciones, a saber:

- Independencia organizativa.
- Independencia conductual.
- Protección de influencia externa.
- Evasión de conflictos de interés.

En el cuadro 1.4 se presentan criterios e indicadores para evaluar la independencia de una organización de evaluación, que provienen de varias fuentes.

Cuadro 1.4 *Criterios e indicadores para establecer la independencia de las organizaciones de evaluación*

Criterio	*Aspecto*	*Indicadores*
Independencia organizativa	La estructura y papel de la unidad de evaluación es apropiada.	Si la unidad de evaluación tiene una declaración de mandato que clarifique que el alcance de su responsabilidad se extiende a todas las operaciones de la organización, y que su línea de reporte de información, personal, presupuesto y funciones es organizativamente independiente de los departamentos operativos, de políticas y estrategia de la organización y de la toma de decisiones relacionada.
	La unidad es responsable y presenta los resultados de la evaluación ante el jefe o jefe encargado de la organización o su junta de gobierno.	Si existe relación de información directa entre la unidad y la gerencia o la junta de la institución.
	La unidad está localizada organizativamente fuera de las funciones de personal o jefe de línea del programa, la actividad o la entidad objeto de la evaluación.	Posición de la unidad en la organización relativa al programa, la actividad o la entidad objeto de la evaluación.
	La unidad informa regularmente al comité superior de auditoría de la organización u otro órgano de supervisión.	Relación de información y frecuencia de presentación de información al órgano supervisor.
	La unidad se encuentra lo suficientemente al margen de presiones políticas para informar sobre los hallazgos sin temor a repercusiones.	Grado al cual la unidad de evaluación y su personal no son responsables ante las autoridades políticas y se encuentran aislados de la participación en actividades políticas.
	El personal de la unidad se encuentra protegido por un sistema de administración de personal en el que la remuneración, la capacitación, el término del cargo y la promoción se basan en el mérito.	Grado al cual el sistema de méritos que cubre la remuneración, capacitación, período del cargo y promoción está vigente y se hace cumplir.
	La unidad tiene acceso a toda la información requerida y a las fuentes de información.	Grado al cual la unidad de evaluación tiene acceso irrestricto al personal, registros, cofinanciadores de la organización y otros asociados y clientes, y los de programas, actividades o entidades que financia o patrocina.
Independencia conductual	La unidad tiene capacidad y disposición de emitir informes fuertes e inflexibles.	Grado al cual la unidad de evaluación ha emitido informes que insten al escrutinio público (dentro de salvaguardias apropiadas

(Continúa en la página siguiente)

(Continuación cuadro 1.4)

Criterio	*Aspecto*	*Indicadores*
		para proteger información confidencial o propietaria y mitigar el riesgo institucional) de las enseñanzas de los programas y actividades de la organización; proponer estándares de desempeño por anticipado de los de uso corriente por parte de la organización; criticar los resultados de los programas, actividades y entidades de la organización.
	La unidad está en capacidad de informar con franqueza.	Grado al cual el mandato de la organización ordena que la unidad de evaluación remita sus informes a la gerencia o junta después de su revisión y comentarios por parte de unidades corporativas pertinentes, pero sin restricciones impuestas por la gerencia sobre su alcance y comentarios.
	La presentación de información de los hallazgos de la evaluación es transparente.	Grado al cual las reglas de divulgación de la organización permitan informar hallazgos significativos a partes interesadas afectadas, internos y externos (dentro de las salvaguardias apropiadas para proteger la información confidencial o propietaria y mitigar el riesgo institucional).
Protección de interferencia externa	La evaluación se diseñó y ejecutó apropiadamente.	Grado al cual la unidad de evaluación tiene capacidad de establecer el diseño, alcance, oportunidad y conducta de las evaluaciones sin interferencia de la gerencia.
	El estudio de evaluación se financió adecuadamente.	Grado al cual la unidad se encuentra libre de restricciones sobre fondos u otros recursos que afecten en forma adversa su capacidad de ejecutar sus responsabilidades.
	No se tiene dominio sobre los juicios del evaluador concernientes al contenido del informe.	Grado al cual el juicio del evaluador en cuanto a lo apropiado del contenido de un informe no se encuentra sujeto a anulación o influencia de autoridad externa.
	Se documentan los procedimientos de recursos humanos independientes para el jefe de la unidad de evaluación.	Grado al cual el mandato o documento equivalente especifica procedimientos para la contratación, el despido, el período de cargos, la revisión de desempeño y la remuneración del jefe de la unidad de evaluación que asegure independencia de la gerencia operativa.
	La unidad tiene control sobre la contratación, la promoción y/o el despido de personal.	Grado al cual la unidad de evaluación tiene control sobre la contratación, la promoción, los incrementos salariales y el despido de personal, dentro de un sistema de méritos.

(Continúa en la página siguiente)

(Continuación cuadro 1.4)

Criterio	*Aspecto*	*Indicadores*
	El empleo permanente del evaluador no se basa en los resultados de la evaluación.	Grado al cual el empleo permanente del evaluador se basa sólo en el desempeño, la competencia y la necesidad de servicios de un evaluador.
Evasión de conflictos de interés	No existen relaciones oficiales, profesionales, personales o financieras que puedan causar que un evaluador limite el grado de una investigación, su divulgación, o debilite o incline los hallazgos.	Grado al cual las políticas y los procedimientos estén vigentes para identificar relaciones del evaluador que puedan interferir con la independencia de la evaluación, se comunican las políticas y los procedimientos al personal por medio de capacitación y de otros medios y se hacen cumplir.
	El evaluador no tiene ideas preconcebidas, prejuicios o sesgos sociales o políticos que puedan afectar los hallazgos de la evaluación.	Grado al cual se encuentran vigentes y se hacen cumplir las políticas y los procedimientos que exigen que los evaluadores valoren e informen prejuicios o predisposiciones personales que arriesguen su capacidad de objetividad para la evaluación, y sobre las que se consulte a partes interesadas como parte del proceso de evaluación para asegurarse contra prejuicios del evaluador.
	El evaluador no participa actualmente ni participó anteriormente en el programa, actividad o entidad objeto de la evaluación a nivel de toma de decisiones, o en algún cargo de gestión financiera o contable y no busca empleo en tal programa, actividad o entidad mientras esté realizando la evaluación.	Grado al cual se presenten y hagan cumplir las reglas o los procedimientos de personal que impidan al personal evaluar programas, actividades o entidades para los cuales tienen o han tenido papel de toma de decisiones o de gestión financiera, o en los cuales estén buscando empleo.
	El evaluador no tiene interés financiero en el programa, la actividad o la entidad objeto de la evaluación.	Grado al cual se encuentren vigentes y se hagan cumplir reglas o procedimientos de personal que impidan al personal evaluar programas, actividades o entidades en las que tienen algún interés financiero.
	Los miembros inmediatos o cercanos de la familia (del evaluador) no participan o están en condición de ejercer influencia directa o significativa sobre el programa, la actividad o la entidad objeto de la evaluación.	Grado al cual se encuentren vigentes y se hagan cumplir reglas o procedimientos de personal que impidan a éstos evaluar programas, actividades o entidades en las que tienen influencia familiares.

Fuente: Danish Ministry of Foreign Affairs, 1999; OCDE, 1991b; CIDA, 2000; Institute of Internal Auditors, 2000; European Federation of Accountants, 2001; INTOSAI, 2001; U.S. GAO, 2002.

001El camino.indd 30 08/11/2010 03:27:41 p.m.

Ejemplos de evaluaciones de desarrollo

Los recuadros 1.4 a 1.8 provienen de informes de evaluaciones (Alnap, 2006) y cada uno es un ejemplo de los criterios descritos anteriormente; pertinencia, eficacia, eficiencia, impacto y sostenibilidad.

Recuadro 1.4 *Pertinencia: La World Food Programme's Evaluation of Food Aid for Relief and Recovery en Somalia*

Esta evaluación fue llevada a cabo por dos expatriados que visitaron Somalia durante tres semanas, a mediados de julio de 2001. En la evaluación se valoró el apoyo de tres años durante los que se distribuyeron 63.000 millones de toneladas de productos básicos alimenticios a 1,3 millones de personas, por un costo de US$55 millones. De este apoyo, el 51% debió ser para rehabilitación y recuperación, el 30% para ayuda de emergencia y el 19% para instituciones sociales. El propósito principal de la ayuda prolongada y la operación de recuperación fue "contribuir al marco general de programas integrados de rehabilitación en Somalia, manteniendo flexibilidad para aprovechar oportunidades de desarrollo y responder a situaciones de emergencia" (WFP, 2002, p. 4). Por consiguiente, en la evaluación se debía examinar la pertinencia de esta combinación de asignaciones, lo mismo que lo apropiado de cada tipo de intervención.

Se consideró la pertinencia general de la intervención en el contexto de la economía política de la ayuda en Somalia. En la evaluación se tuvo en cuenta el fundamento para proporcionar ayuda alimentaria a Somalia. Los argumentos en contra de esta ayuda incluyeron el hecho de que Somalia se encuentra usualmente en déficit alimenticio, la población en muchas ubicaciones se encuentra aislada de los mercados convencionales, y muchos somalíes perdieron su ocupación primaria y sus activos. Los argumentos en contra de la ayuda alimenticia sugerían que podría tener más sentido dar a los beneficiarios fondos con los cuales adquirir alimentos locales en donde estuviesen disponibles, ya fuera en forma de dinero para trabajo o de alimento para trabajo. Sin embargo, tales compromisos suelen ser proyectos de largo plazo, sin una estrategia clara de salida. El examen de esta evaluación, tanto de los temas generales como de los específicos, significa que su análisis de pertinencia es integral.

Recuadro 1.5 *Eficacia: evaluación del apoyo del DfID para los World Food Programme's Efforts en Bangladesh*

En septiembre de 2000, las inundaciones en seis distritos de Bangladesh afectaron gravemente a unos 2,7 millones de personas. El DfID apoyó el World Food Programme, aportando tres distribuciones de alimentos, incluyendo una ración completa de arroz, legumbres y aceite. En la primera distribución, 260.000 personas recibieron apoyo alimentario; en la segunda y tercera distribuciones, lo recibieron 420.000 beneficiarios. La evaluación del DfID (DfID, 2001) aportó un análisis integral sobre si se cumplieron los objetivos del proyecto con respecto al tamaño de las raciones, la mezcla de productos y el programa de distribuciones.

En la evaluación se emplearon métodos cuantitativos y cualitativos. Los datos cuantitativos se recopilaron en 2.644 hogares seleccionados al azar en aldeas de la zona del proyecto, y los cualitativos se recopilaron durante evaluaciones sobre medios de vida en seis aldeas representativas sobre los sistemas, condiciones y perspectivas referentes a los medios de vida en comunidades afectadas por inundaciones. Unas cinco semanas después se envió un segundo equipo más pequeño de evaluación, después de la primera evaluación cualitativa, con el fin de explorar percepciones y comportamientos comunitarios relacionados con la ración alimentaria, incluyendo temas como la oportunidad de la distribución, la conveniencia de los productos suministrados y los patrones de uso. Los conjuntos de datos cuantitativos y cualitativos se utilizaron en forma combinada en el análisis.

El informe incluye la mayoría de los elementos básicos para la evaluación de la eficacia, entre ellos:

- Un examen para el desarrollo de los objetivos de la intervención, incluyendo un análisis del marco lógico.
- Evaluación de los criterios utilizados para la selección de beneficiarios, incluyendo las opiniones de partes interesadas primarias sobre estos criterios.
- Análisis de los mecanismos de implementación, incluyendo niveles de participación comunitaria.
- Una estimación de la precisión de la focalización, desagregada por géneros y grupos socioeconómicos.
- Evaluación de los recursos provistos (tamaño de las raciones y mezcla de productos), incluyendo las razones para su provisión (este campo puede evaluarse también con el criterio de pertinencia).
- Examen de la suficiencia de los programas de distribución.
- Análisis de las opiniones de los beneficiarios de la intervención.

Recuadro 1.6 *Eficiencia: evaluación de los Disasters Emergency Committee's Mozambique Flood Appeal Funds*

Después de las inundaciones del año 2000 en Mozambique, el grupo de evaluación del Disasters Emergency Committee (DEC) examinó la respuesta humanitaria emprendida por las agencias del DEC (DEC, 2001). El propósito de la evaluación consistía en informar al público británico sobre cómo y dónde se utilizaron sus fondos, e identificar buenas prácticas para futuras operaciones de emergencia. El método para la evaluación incluyó extensas entrevistas, investigación de antecedentes, visitas de campo y una encuesta detallada a los beneficiarios.

El capítulo dedicado a la eficiencia contiene muchos de los elementos básicos necesarios para la evaluación, e incluye el análisis de:

- El uso de activos militares por parte de las agencias del DEC, evaluado en términos de falta de colaboración en el uso de los helicópteros para realizar la evaluación de las necesidades; los altos costos de utilizar fuerzas militares occidentales en lugar de instalaciones comerciales para ayuda humanitaria; y los costos comparativos de la Royal Air Force, el Ejército de EUA y las South African National Defence Forces (en el informe se anota que costosas operaciones militares consumieron grandes cantidades de fondos, lo que limitó la posterior financiación de donantes para proyectos de ONG).
- Los efectos en la eficiencia de un mercado subdesarrollado para la contratación de servicios (por ejemplo, aunque el uso de contratistas nacionales permitió que las agencias ejecutaran obras de equipo pesado, como reparación de vías, sin necesidad de efectuar grandes inversiones de capital, los contratistas que utilizaron las agencias del DEC incumplieron a menudo sus obligaciones en forma oportuna).
- La eficiencia de la elección de una respuesta (intervenir directamente con programas operativos, trabajar con asociados locales, o con redes internacionales); el grupo de evaluación encontró que la composición del personal constituyó un factor más importante para establecer la eficiencia que la elección de una respuesta (este campo también podría examinarse con el criterio de pertinencia).
- Si fue más eficiente que las agencias construyeran su respuesta con base en la capacidad existente en el país o con personal internacional.
- Si las agencias con asociados existentes fueron más eficientes que aquellas que no contaron con dichos asociados.
- Cómo la inversión en el alistamiento condujo a una respuesta más eficiente.
- La eficiencia de los sistemas contables.

Se realizó un intento para comparar los costos de los insumos entre las distintas agencias, pero resultó imposible efectuar dicha comparación debido a las diferencias en los artículos provistos y los canales de entrega utilizados. Así, la evaluación se basó en las implicaciones generales de los costos de las prácticas realizadas, como los costos de almacenamiento y transporte. La evaluación incluyó también un desglose del empleo de fondos por sectores y, para cada agencia del DEC, por suministros y materiales, costos de personal, otros costos y costos de gerencia de la agencia.

Recuadro 1.7 *Impacto: Joint Evaluation of Emergency Assistance to Rwanda*

La Joint Evaluation of Emergency Assistance to Rwanda (JEEAR, 1996) constituye la evaluación más grande y completa de acción humanitaria que se haya realizado. Incluyó a 52 consultores e investigadores y en el informe se fijaron estándares para la evaluación conjunta del impacto de la acción política (y la falta de ésta) en emergencias complejas.

Con JEEAR se evaluó el impacto principalmente en términos de una falta de intervención en Ruanda por parte de la comunidad internacional, a pesar de señales significativas de que las fuerzas militares de ese país preparaban el clima y las estructuras para el genocidio y el asesinato político. Empleó una definición de acción humanitaria que incluyó funciones políticas y socioeconómicas, definición que llevó a un análisis de las estructuras políticas que determinan en gran medida la respuesta y el impacto humanitario.

La falta de intervención se consideró en dos partes: un análisis de los factores históricos que explicaron el genocidio y una descripción detallada de los eventos inmediatos que ocasionaron el genocidio. El valor de la evaluación conjunta es que fue más allá de los límites del examen de las intervenciones de un solo sector, a un análisis de economía política. El enfoque de economía política fue entonces vinculado a la evaluación de la eficacia de la respuesta humanitaria.

Este enfoque puede contrastarse con el utilizado en evaluaciones de otras crisis, como el conflicto de Kosovo y sus secuelas, los efectos del huracán Mitch y las intervenciones en Afganistán. En cada uno de estos casos, se tomaron decisiones para realizar evaluaciones de una sola agencia y un solo sector, que en gran parte fracasaron al no capturar la naturaleza política del evento y la respuesta a éste. En los casos de Kosovo y Afganistán, esto condujo a una falta de atención a los problemas de protección y derechos humanos por parte de los evaluadores (ALNAP, 2001, 2004). En el caso del huracán Mitch, condujo a una falta de atención sobre hasta qué punto la acción humanitaria apoyó la agenda de transformación propuesta en la Declaración de Estocolmo (ALNAP, 2002).

La JEEAR es inusual en su evaluación de impacto, porque hace mucho hincapié en por qué hubo tan poco interés en intervenir en Ruanda (principalmente por su falta de significación geopolítica), en lugar de relacionar los eventos y sus consecuencias. Una de las enseñanzas para los evaluadores es que las evaluaciones de impacto deben trabajar no sólo en qué intervenciones se realizaron, sino también en qué pudo haber sucedido dadas otras circunstancias y diferentes clases de intervención.

Recuadro 1.8 *Sostenibilidad: evaluación de la JICA del Third Country Training Program on Information and Communication Technology*

La Agencia Japonesa de Cooperación Internacional (JICA) realizó una evaluación de un proyecto en las Filipinas, dirigido a ofrecer una oportunidad a participantes de Camboya, Laos RDP, Myanmar y Vietnam, con el fin de mejorar su conocimiento y técnicas en el campo de la tecnología de información y comunicaciones para la capacidad empresarial.

En la evaluación (JICA, 2005b) se concluyó que la sostenibilidad era alta, dado el compromiso de la Foundation for Information Technology and Education Development (FIT-ED) de asumir futuros programas de capacitación para lograr los objetivos del proyecto. La FIT-ED instituyó un *e-group* (grupo de comunicación electrónica) con el fin de permitir el trabajo en red de los participantes, el conocimiento compartido de la FIT-ED y enriquecer sus capacidades. Como institución comprometida a incrementar la conciencia sobre la tecnología de información en los sectores del gobierno y empresarial en los países de la Association of Southeast Asian Nations (ASEAN), FIT-ED continuará a la vanguardia de las actividades de la asociación relacionadas con la tecnología de información y comunicaciones.

La asignación adecuada y oportuna de recursos de FIT-ED para los tres cursos de capacitación demostró su compromiso de sostener el programa de capacitación. Los participantes expresaron también su compromiso de apoyar la iniciativa y reconocieron la importancia de la tecnología de información y comunicaciones para sus negocios; el 84% de los entrevistados ya estaba aplicando conocimiento y destrezas adquiridas en el programa de capacitación a su trabajo (en desarrollo de sitios web, comunicaciones, textiles y vestuario, importación y exportación de artesanías, construcción, producción de café e iniciativas del gobierno, entre otros campos). Los entrevistados informaron haberse beneficiado mucho con el curso, al que consideraron como el inicio del programa de capacitación. Además de utilizar el plan estratégico de negocios electrónicos esbozado durante el programa de capacitación como referencia, los participantes también emplearon la Internet para aplicar el conocimiento obtenido en el curso para promover los sectores ya citados.

Resumen

La evaluación se ha realizado desde hace siglos, aunque sólo recientemente se ha encargado de los efectos de las intervenciones para el desarrollo.

La evaluación es de tres formas (formativa, final y prospectiva) y tiene cuatro propósitos (ético, gerencial, de decisiones y educativo y motivacional). Puede suministrar información sobre estrategia (¿se están haciendo las cosas correctas?), operaciones (¿se están haciendo bien?) y aprendizaje (¿hay mejores maneras de hacerlas?). La evaluación puede efectuarse en forma interna, externa o participativa.

La evaluación para el desarrollo evolucionó a partir de la investigación de las ciencias sociales, el método científico y la auditoría. El papel del evaluador ha cambiado a través del tiempo, de un énfasis en el evaluador como auditor, contador y certificador, a un énfasis como investigador y facilitador de evaluaciones participativas.

La evaluación para el desarrollo se basa en los criterios de la OCDE/CAD de pertinencia, eficacia, eficiencia, impacto y sostenibilidad. La OCDE/CAD desarrolló también principios específicos para la evaluación de la ayuda para el desarrollo y estándares de calidad de evaluación.

Una parte importante de la credibilidad es su independencia. Los directores de los bancos de desarrollo multilateral identificaron cuatro dimensiones o criterios de independencia de evaluación: independencia organizativa, conductual, evasión de conflictos de interés y protección de interferencia externa.

Actividad del capítulo 1

Ejercicios de aplicación 1.1

1. Suponga que se le ha solicitado justificar por qué la evaluación para el desarrollo debe constituir un gasto en el presupuesto para un nuevo programa nacional. El programa se diseñó con el fin de mejorar la educación de las familias sobre prácticas eficaces para la salud. ¿Qué diría usted en defensa de la educación para el desarrollo?
2. Entreviste a un evaluador en su campo, a fin de establecer el grado al cual se aplican los estándares y principios orientadores en las evaluaciones que él o ella ha visto. (Si no tiene acceso a un evaluador, revise informes recientes de evaluaciones efectuadas en su campo.) ¿Dónde parecen estar las fortalezas? ¿Dónde las debilidades? Comparta sus hallazgos con colegas de evaluación y escuche sus comentarios y experiencias. ¿Encuentra algunos patrones?

Referencias y lecturas adicionales

Adamo, Abra. 2003. *Mainstreaming Gender in IDRC's MINGA Program Initiative: A Formative Evaluation*. https://idl-bnc.idrc.ca/dspace/bitstream/123456789/30972/1/121759.pdf.

ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action). 2001. *Humanitarian Action: Learning from Evaluation*. ALNAP. *Annual Review 2001*. ALNAP/Overseas Development Institute, Londres.

———. 2002. *Humanitarian Action: Improved Performance through Improved Learning. ALNAP Annual Review 2002*. ALNAP/Overseas Development Institute, Londres.

———. 2006. *Evaluating Humanitarian Action Using the OECD/DAC Criteria*. Overseas Development Institute, Londres.
http://www.odi.org.uk/alnap/publications/eha_dac/pdfs/eha_2006.pdf.

Banco Asiático de Desarrollo (BAsD). 2007. *Mongolia: Second Financial Sector Program*.
http://www.oecd.org/dataoecd/59/35/39926954.pdf.

Banco Interamericano de Desarrollo. 2004. *Proposal for Sector-wide Approaches*.
http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=509733.

Bhy, Y. Tan Sri Data' Setia Ambrin bin Huang. 2006. *The Role of the National Audit Department of Malaysia in Promoting Government Accountability*. Documento presentado en el Third Symposium of the Asian Organisation of Supreme audit. Institutions (Asosai), Shanghai, septiembre 13. http://apps.emoe.gov.my/bad/NADRole.htm.

Brooks, R. A. 1996. "Blending Two Cultures: State Legislative Auditing and Evaluation". En *Evaluation and Auditing: Prospects for Convergence,* ed. Carl Wisler, 15-28. *New Directions for Evaluation* 71 (otoño). San Francisco: Jossey-Bass.

Burkhead, J. 1956. *Government Budgeting*. Nueva York: John Wiley and Sons.

Callow-Heusser, Catherine. 2002. *Digital Resources for Evaluators*. http://www.resources4evaluators.info/CommunitiesOfEvaluators.html.

Chelimsky, Eleanor. 1995. "Preamble: New Dimensions in Evaluation". En *Evaluating Country Development Policies and Programs,* ed. Robert Picciotto y Ray C. Rist, 3-8. *New Approaches for a New Agenda* 67 (otoño). Publicación de la American Evaluation Association. San Francisco: Jossey-Bass Publishers.

———. 1997. "The Coming Transformations in Evaluation". En *Evaluation for the 21st Century: A Handbook*, eds. E. Chelimsky y W. R. Shadish, pp. 1-26. Thousand Oaks, CA: Sage Publications.

Chelimsky, Eleanor y William R. Shadish. 1997. *Evaluation for the 21st Century: A Handbook.* Thousand Oaks, CA: Sage Publications.

CIDA (Canadian International Development Agency). 2000. *CIDA Evaluation Guide*. Ottawa.

DANIDA, Ministry of Foreign Affairs. 1998, 1999. Guidelines for an Output and Outcome Indicator System. Copenhague: DANIDA.

DEC (Disasters Emergency Committee). 2001. *Independent Evaluation of DEC Mozambique Floods Appeal Funds: March 2000-December 2000*. Londres.

Derlien, Hans-Ulrich. 1999. "Program Evaluation in the Federal Republic of Germany". En *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations*, ed. Ray C. Rist, 37-52. Nueva Brunswick, NJ: Transaction Publishers.

DfID (Department for International Development). *Sector Wide Approaches (SWAps)*. Londres. http://www.dfid.gov.uk/mdg/aid-effectiveness/swaps.asp.

———. 2001. *Emergency Food Aid to Flood-Affected People in South-Western Bangladesh: Evaluation report*. Londres.

European Federation of Accountants. 2001. *The Conceptual Approach to Protecting Auditor Independence*. Bruxelles.

Evaluation Cooperation Group of the Multilateral Development Banks. s.f. *Good Practice Standards for Evaluation of MDB–Supported Public Sector Operations*. Working Group on Evaluation Criteria and Ratings for Public Sector Evaluation. https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_B084A3976FF5F808482571D90027AD16/1E8F8A3670331832482574630002F0726/

Feuerstein, M. T. 1986. *Partners in Evaluation: Evaluating Development and Community Programs with Participants*. Londres: MacMillan, en asociación con Teaching Aids at Low Cost.

Fitzpatrick, Jody L., James, R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* 3ª ed. Nueva York: Pearson Education, Inc.

Fontaine, C. y E. Monnier. 2002. "Evaluation in France". En *International Atlas of Evaluation,* eds. Jan-Eric Furubo, Ray C. Rist y Rolf Sandahl, 63-76. Nueva Brunswick, NJ: Transaction Publishers.

Furubo, Jan-Eric, Ray C. Rist y Rolf Sandahl, eds. 2002. *International Atlas of Evaluation*. Nueva Brunswick, NJ: Transaction Publishers.

Furubo, Jan-Eric y R. Sandahl. 2002. "Coordinated Pluralism". En *International Atlas of Evaluation*, eds. Jan-Eric Furubo, Ray C. Rist y Rolf Sandahl, 115-28. Nueva Brunswick, NJ: Transaction Publishers.

Gobierno de Malawi. 2007. "National Indicators for Routine Monitoring of Quality of Health Services at Central Hospital". http://www.malawi.gov.mw/MoHP/Information/Central%20Hospital%20Indicators.htm.

Human Rights Education Associated. 1997. *Evaluation in the Human Rights Education Field: Getting Started.* Netherlands Helsinki Committee. http://www.hrea.org/pubs/EvaluationGuide/.

Insideout. 2005. "M&E In's and Out's". *Insideout* 3 (octubre/noviembre) p. 1. http://www.insideoutresearch.co.za/news_l/Newsletter_issue%203.pdf.

Institute of Internal Auditors. 2000. *Professional Practices Framework*. Altamente Springs, Florida.

INTOSAI (Organización Internacional de Entidades Fiscalizadoras Superiores). s.f.. *Draft Strategic Plan 2004 to 2009*. Viena, Austria. http://www.gao.gov/cghome/parwi/img4.html.

———. 2001. *Code of Ethics and Auditing Standards*. Estocolmo.

JEEAR (Joint Evaluation of Emergency Assistance to Rwanda). 1996. *The International Response to Conflict and Genocide: Lessons from the Rwanda Experience, 5 volumes*. JEEAR Steering Committee, Copenhague.

JICA (Agencia Japonesa de Cooperación Internacional). 2005a. *JICA Evaluation: Information and Communication*. Tokyo. http://www.jica.go.jp/english/operations/evaluation/.

———. 2005b. *Results of Evaluation, Achievement of the Project Joint Evaluation of Emergency Assistance to Rwanda*. Tokio.

KRA Corporation. 1997. *A Guide to Evaluating Crime Control of Programs in Public Housing*. Informe preparado para el U.S. Department of Housing and Urban Development. http://www.ojp.usdoj.gov/BJA/evaluation/guide/documents/guide_to_evaluating_crime.html

Kusek, Jody Zall y Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Banco Mundial, Washington, DC. http://www.oecd.org/dataoecd/23/27/35281194.pdf.

Lawrence, J. 1989. "Engaging Recipients in Development Evaluation: The 'Stakeholder' Approach". *Evaluation Review* 13 (3): 243-56.

Measure Evaluation. s.f.. *Monitoring and Evaluation of Population and Health Programs*. University of North Carolina, Chapel Hill. http://www.cpc.unc.edu/measure.

Mikesell, J. L. 1995. *Fiscal Administration: Analysis and Applications for the Public Sector*, 4ª ed. Belmont, CA: Wadsworth Publishing Company.

Molund, Stefan y Göran Schill 2004. *Looking Back, Moving Forward: SIDA Evaluation Manual*. Swedish International Development Agency, Estocolmo.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 1991a. *DAC Criteria for Evaluating Development Assistance*. Development Assistance Committee. http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html.

———. 1991b. *Principles for Evaluation of Development Assistance*. Development Assistance Committee. http://siteresources.worldbank.org/EXTGLOREGPARPRO/Resources/DACPrinciples1991.pdf.

———. 1998. *Review of the DAC Principles for Evaluation of Development Assistance*. Development Assistance Committee. http://www.oecd.org/dataoecd/31/12/2755284.pdf.

———. 2002. *OECD Glossary of Key Terms in Evaluation and Results-Based Management*. Development Assistance Committee, París.

———. 2006. *Evaluation Quality Standards for Test Phase Application*. Development Assistance Committee. http://www.oecd.org/dataoecd/30/62/36596604.pdf.

———. 2007a. General information. http://www.oecd.org/document/48/0,3343,en_2649_201185_1876912_1_1_1_1, 00.html.

———. 2007b. Development Co-operation Directorate DCD-DAC. http://www.oecd.org/department/0,2688,en_2649_33721_1_1_1_1_1,00.html.

Office of the Secretary of Defense Comptroller Center. 2007. *The Historical Context*. http://www.defenselink.mil/comptroller/icenter/budget/histcontext.htm.

Picciotto, Robert. 2002. "Development Evaluation as a Discipline". International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio.

———. 2005. "The Value of Evaluation Standards: A Comparative Assessment". *Journal of Multidisciplinary Evaluation* 3: 30-59. http://evaluation.wmich.edu/jmde/content/JMDE003content/PDFspercent20JMDEpercent20003/4_percent20The_Value_of_Evaluation_Standards_A_Comparative_Assessment.pdf.

Quesnel, Jean Serge. 2006. "The Importance of Evaluation Associations and Networks". En *New Trends in Development Evaluation* 5, UNICEF Regional Office for Central and Eastern Europe, Commonwealth of Independent States, and International Program Evaluation Network. http://www.unicef.org/ceecis/New_trends_Dev_EValuation.pdf.

Quinn, Michael 1997. *Utilization-Focused Evaluation.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Rossi, Peter y Howard Freeman. 1993. *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage Publications.

Scientific and National Councils for Evaluation. 1999. *A Practical Guide to Program and Policy Evaluation, París:* National Council of Evaluation.

Sonnichsen, R. C. 2000. *High-Impact Internal Evaluation.* Thousand Oaks, CA: Sage Publications.

Treasury Board of Canada Secretariat. 1993. "Linkages between Audit and Evaluation in Canadian Federal Developments". http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/TB_h4/evaluation03_e.asp.

Tyler, C. y J. Willand. 1997. "Public Budgeting in America: A Twentieth Century Retrospective". *Journal of Public Budgeting, Accounting and Financial Management* 9 (2): 189-219. http://www.ipspr.sc.edu/publication/Budgeting_in_America.htm.

U.S. Department of Housing and Urban Development. 1997. "A Guide to Evaluating Crime Control of Programs in Public Housing". Documento preparado por la KRA Corporation, Washington, DC. http://www.ojp.usdoj.gov/BJA/evaluation/guide/documents/benefits_of_evaluation.htm.

U.S. GAO (General Accounting Office). 1986. *Teenage Pregnancy: 500,000 Births a Year but Few Tested Programs*. Washington, DC: U.S. GAO.

———. 2002. "Government Auditing Standards, Amendment 3", GAO, Washington, DC.

Weiss, Carol. 2004. "Identifying the Intended Use(s) of an Evaluation". *Evaluation Guideline 6.* http://www.idrc.ca/ev_en.php?ID=58213_201&ID2=DO_TOPIC p 1.

WFP (World Food Programme). 2002. *Full Report of the Evaluation of PRRO Somalia 6073.00, Food Aid for Relief and Recovery in Somalia*. Roma.

Sitios web

Organizaciones de evaluación

African Evaluation Association. http://www.afrea.org/.
American Evaluation Association. http://www.eval.org.
Australasian Evaluation Society. http://www.aes.asn.au.
Brazilian Evaluation Association. http://www.avaliabrasil.org.br.
Canadian Evaluation Society. http://www.evaluationcanada.ca.
Danish Evaluation Society. http://www.danskevalueringsselskab.dk.
Dutch Evaluation Society. http://www.videnet.nl/.
European Evaluation Society. http://www.europeanevaluation.org.
Finnish Evaluation Society. http://www.finnishevaluationsociety.net/.
French Evaluation Society. http://www.sfe.asso.fr/.
German Society for Evaluation Standards. http://www.degeval.de/.

Institute of Internal Auditors. http://www.theiia.org.
International Development Evaluation Association (IDEAS). http://www.ideas-int.org/. (IDEAS página web con vínculos a muchas organizaciones: http://www.ideas-int.org/Links.aspx.)
International Organisation for Cooperation in Evaluation. http://ioce.net/overview/general.shtml.
International Organization of Supreme Audit Institutions (INTOSAI). http://www.intosai.org/.
International Program Evaluation Network (Rusia y los recientes estados independientes). http://www.eval-net.org/.
Israeli Association for Program Evaluation. http://www.iape.org.il.
Italian Evaluation Association. http://www.valutazioneitaliana.it/
Japan Evaluation Society. http://ioce.net/members/eval_associations.shtml
Latin American and Caribbean Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects (Preval). http://www.preval.org/.
Malaysian Evaluation Society. http://www.mes.org.my.
Nigerian Network of Monitoring and Evaluation. http://www.pnud.ne/rense/.
Polish Evaluation Society. http://www.pte.org.pl/x.php/1,71/Strona-glowna.html.
Quebecois Society for Program Evaluation. http://www.sqep.ca.
Red de Evaluación de América Latina y el Caribe. http://www.relacweb.org.
South African Evaluation Network. http://www.afrea.org/webs/southafrica/.
South African Monitoring and Evaluation Association. http://www.samea.org.za.
Spanish Evaluation Society. http://www.sociedadevaluacion.org.
Sri Lankan Evaluation Association. http://www.nsf.ac.lk/sleva/.
Swedish Evaluation Society. http://www.svuf.nu.
Swiss Evaluation Society. http://www.seval.ch/de/index.cfm.
Ugandan Evaluation Association. http://www.ueas.org.
United Kingdom Evaluation Society. http://www.evaluation.org.uk.
Wallonian Society for Evaluation. http://www.prospeval.org.

Estándares de evaluación

African Evaluation Association. *Evaluation Standards and Guidelines*. http://www.afrea.org/.
American Evaluation Association. *Guiding Principles*. http://www.eval.org/Publications/GuidingPrinciples.asp.
Australasian Evaluation Society. *Ethical Guidelines for Evaluators*. http://www.aes.asn.au/content/ethics_guidelines.pdf.
ECGnet [The Evaluation Cooperation Group] https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_B084A3976FF5F808482571D90027AD16/1E8F8A3670331832482574630O2F0726/.
FNUAP (Fondo de Población de las Naciones Unidas). s.f. *List of Evaluation Reports and Findings*. http://www.unfpa.org/publications/index.cfm.
German Society for Evaluation. *Standards*. http://www.degeval.de/standards/standards.htm.
Italian Evaluation Association. *Guidelines*. http://www.valutazioneitaliana.it/statuto.htm#Linee.
OCDE (Organización para la Cooperación y el Desarrollo Económicos). *Evaluation Quality Standards* (para aplicación de fase de prueba). Development Assistance Committee. http://www.oecd.org/dataoecd/30/62/36596604.pdf.
———. *Evaluating Development Co-Operation: Summary of Key Norms and Standards.* Development Assistance Committee, Network on Development Evaluation. http://www.oecd.org/dac/evaluationnetwork.
Program Evaluation Standards (actualización 1998). http://www.eval.org/EvaluationDocuments/progeval.html.

Swiss Evaluation Society. *Standards*. seval.ch/.
UNEG (Grupo de Evaluación de las Naciones Unidas). *Norms for Evaluation in the UN System*. http://www.uneval.org/docs/ACFFC9F.pdf.
———. *Standards for Evaluation in the UN System*. http://www.uneval.org/docs/ACFFCA1.pdf.
UNICEF Regional Office for Central and Eastern Europe, Commonwealth of Independent States, and International Program Evaluation Network. 2006. *New Trends in Development Evaluation* 5. http://www.unicef.org/ceecis/New_trends_Dev_EValuation.pdf.
United Nations Development Project Evaluation Office. http://www.undp.org/eo/.

CAPÍTULO 2

Conocimiento de los temas que orientan la evaluación para el desarrollo

El campo de la evaluación para el desarrollo es un campo relativamente nuevo que cambia como respuesta a temas emergentes, en los países desarrollados y en desarrollo. En este capítulo se tratan algunos de los temas actuales que afectan a esos países.

Panorama general de la evaluación en países desarrollados y en desarrollo

La evaluación puede ayudar a los países a establecer qué tan bien, y hasta qué punto, están logrando sus objetivos de desarrollo, incluyendo los Objetivos de Desarrollo del Milenio (ODM). Los gobiernos y otros pueden utilizar las percepciones y recomendaciones básicas de los estudios de evaluación para iniciar el cambio. La evaluación permite que los países utilicen la experiencia para mejorar el diseño y la ejecución de proyectos, programas y políticas, cambiar las direcciones futuras, o ambas cosas.

Muchos países desarrollados y en desarrollo han implementado sistemas de seguimiento y evaluación (SyE) para asistir en el desarrollo. Estos sistemas pueden configurarse en distintas formas, dependiendo de las necesidades y de los recursos disponibles.

001El camino.indd 43 08/11/2010 03:27:41 p.m.

La evaluación en los países desarrollados

La mayoría de los 30 países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) tienen ya sistemas maduros de SyE, aunque su puesta en funcionamiento no fue una tarea fácil y tampoco un proceso lineal. Estos países difieren, a menudo en forma sustancial, en su trayectoria, enfoque, estilo y nivel de desarrollo.

Furubo, Rist y Sandahl (2002) mapearon las culturas de evaluación en los países de la OCDE a fin de explicar los patrones observados, y examinaron nueve variables en 23 países, dándole a cada país un puntaje entre 0 (bajo) y 2 (alto) a cada una de las siguientes variables:

1. La evaluación se realiza en muchos dominios de la política.
2. Existe oferta de evaluadores especializados en distintas disciplinas que dominan diferentes métodos de evaluación y realizan evaluaciones.
3. Existe un discurso nacional referente a la evaluación en el que las discusiones generales se ajustan al entorno nacional específico.
4. Existe una profesión con sociedades propias o con frecuente asistencia a reuniones de sociedades internacionales y al menos alguna discusión referente a las normas y la ética de la profesión.
5. Existen acuerdos institucionales en el gobierno para hacer evaluaciones y difundir sus resultados para conocimiento de quienes toman las decisiones.
6. Existen acuerdos institucionales en el Parlamento o en otros organismos legislativos para hacer evaluaciones y difundirlas para conocimiento de quienes toman las decisiones.
7. Existe un elemento de pluralismo; es decir, dentro de cada dominio de política existen distintas personas o agencias que encargan y hacen evaluaciones.
8. Las actividades de evaluación se realizan también dentro de la máxima autoridad de auditoría.
9. Las evaluaciones no deberían centrarse solamente en la producción técnica o la relación entre insumos y productos. Algunas evaluaciones del sector público deben tener como propósito los efectos directos del programa o de políticas (Furubo, Rist y Sandahl, 2002).

Según estos criterios, Australia, Canadá, los Países Bajos, Suecia y los Estados Unidos obtuvieron los "puntajes de cultura de evaluación" más altos entre los países de la OCDE, en 2002.

Los países de la OCDE desarrollaron su cultura de evaluación y sistemas de SyE como respuesta a distintos grados de presión interna y externa. Por ejemplo, Francia, Alemania y los Países Bajos desarrollaron su cultura de evaluación como respuesta a fuertes presiones tanto internas como externas (principalmente estas últimas relacionadas con la Unión Europea, UE). En contraste, algunos países, como Australia, Canadá, la República de Corea y los Estados Unidos, se vieron motivados más por presiones internas (Furubo, Rist y Sandahl, 2002).

001El camino.indd 44 08/11/2010 03:27:41 p.m.

La primera ola de países de la OCDE tuvo la motivación de adoptar cultura de evaluación, sobre todo por fuertes presiones internas, como la planificación interna, la programación y los imperativos del presupuesto para nuevos programas de gasto socioeconómico, y también por la supervisión legislativa. Varios factores contribuyeron a que se adoptara una cultura de evaluación en los países pioneros. Muchos de los primeros que la adoptaron se encontraban predispuestos a hacerlo por tener un sistema político democrático, fuerte tradición empírica, funcionarios públicos capacitados en ciencias sociales (y no capacitación estrictamente legal), y sistemas administrativos e instituciones eficientes.

Los países con alto nivel de gastos en educación, salud y bienestar social adoptaron mecanismos de evaluación en estos campos, que luego se trasladaron a otros campos de la política pública. Los países de la OCDE que adoptaron temprano una cultura de evaluación fueron también instrumentales para propagar la cultura de evaluación a otros países, difundiendo ideas e información de la evaluación, e iniciando organizaciones de evaluación, institutos de capacitación, redes y empresas de consultoría.

En contraste, muchos de los países recién llegados a la OCDE (entre ellos Irlanda, Italia y España) tuvieron tendencia a responder a los temas de evaluación, principalmente a causa de presiones externas fuertes, sobre todo requisitos de integración a la UE, incluido el acceso a fondos de desarrollo estructural de la Unión. Estos recién llegados recibieron fuerte influencia de la cultura de evaluación de los países de la primera ola, como también de la cultura de evaluación arraigada en las organizaciones internacionales con las cuales interactúan.

El Instituto Travistock (2003) describe un modelo, o mapa, para una jornada hacia el desarrollo de capacidad de evaluación que consta de cuatro etapas y destinos intermedios, a saber:

- Etapa 1: orden de evaluar.
- Etapa 2: coordinar la evaluación.
- Etapa 3: institucionalizar la evaluación.
- Etapa 4: construir un sistema de evaluación.

La etapa 1 suele empezar por la presión externa que ordena la evaluación mediante normas, regulaciones u objetivos de política. Aun cuando la fuerza motora provenga desde el interior, es probable un cierto grado de escrutinio externo.

La etapa 2 incluye dos clases de acciones para responder a la política de evaluación basada en reglas de la primera etapa. La primera proporciona lineamientos y herramientas básicas; la segunda hace hincapié en la profesionalización del personal como forma de mejorar la calidad.

La etapa 3 suele iniciarse después de haberse configurado una unidad central y estar en funcionamiento. Incluye dos pasos, usualmente adoptados de manera simultánea: la creación de unidades descentralizadas y el mejoramiento de la oferta de pericia de evaluación.

001El camino.indd 45 08/11/2010 03:27:41 p.m.

La etapa 4 incluye la construcción de un sistema de evaluación totalmente operativo, en el que se incorporan la evaluación a la formulación de políticas, la administración de los programas y la gobernanza. Incluye el establecimiento de vínculos internos más fuertes dentro del sistema, y la apertura de la red a partes interesadas externas.

Los países pioneros y recién llegados a la OCDE adoptaron un enfoque distinto para crear sistemas de seguimiento y evaluación. Adoptaron uno de estos tres enfoques:

- Enfoque que abarca la totalidad del gobierno.
- Enfoque circunscrito.
- Enfoque mixto.

El enfoque que abarca la totalidad del gobierno

Este enfoque fue adoptado por algunos de los primeros países que acogieron el SyE, como Australia. El enfoque incluye un establecimiento de SyE integral de base amplia a través de todo el gobierno. Un marco de enfoque que abarca la totalidad del gobierno no puede desarrollarse de la noche a la mañana, y puede pasar una década antes de que dicho marco se incorpore de una manera sostenible (Banco Mundial, 1999).

Enfoque que abarca la totalidad del gobierno: Establecimiento de seguimiento y evaluación de base amplia e integral en todas las dependencias gubernamentales.

Para garantizar el éxito de tal enfoque, debe obtenerse y mantenerse el apoyo del gobierno, desarrollarse las habilidades necesarias y configurarse las estructuras y sistemas del servicio civil para utilizar plenamente los hallazgos del SyE. Los países en desarrollo deben también asegurar un apoyo continuo de las agencias de ayuda para el desarrollo.

Con la adopción de los ODM, muchos países en desarrollo buscan diseñar e implementar sistemas de evaluación integrales con el enfoque que abarca la totalidad del gobierno. Con el mayor énfasis en los resultados en los préstamos de ayuda internacional, un mayor número de donantes, gobiernos y otras instituciones está apoyando a los países en desarrollo para ayudarles a construir capacidad y sistemas de evaluación.

El enfoque de Australia para la evaluación. Australia fue un país pionero en el desarrollo de sistemas de seguimiento y evaluación, al haberlos iniciado ya desde 1987. Las ventajas intrínsecas conducentes a una cultura y estructura sólida de evaluación incluyeron las siguientes:

- Una fuerte capacidad humana, institucional y gerencial en el sector público.
- Un servicio público reconocido por su integridad, honestidad y profesionalismo.
- Sistemas financieros, presupuestarios y contables bien desarrollados.
- Una tradición de rendición de cuentas y transparencia.
- Dirigentes políticos dignos de crédito y legítimos.

Dos factores principales contribuyeron al éxito de los sólidos sistemas de seguimiento y evaluación australianos. Primero, las restricciones presupuestarias incitaron al gobierno

a buscar maneras de obtener un mayor valor por el dinero. Segundo, Australia contó también con dos dirigentes institucionales importantes para la evaluación: el Departamento de Finanzas y la Oficina Nacional de Auditoría australiana. Además, tuvo el apoyo de los miembros del gabinete y ministros clave, que le dieron importancia a utilizar los hallazgos de la evaluación para mejorar la toma de decisiones informada (Mackay, 2002).

La primera generación de evaluación (1987-97) empezó en un tiempo de fuertes presiones en el presupuesto. Muchas reformas del sector público implicaron la autonomía de departamentos y agencias de línea, pero omitieron realizar seguimiento y evaluación, por lo que los gobiernos obligaron a departamentos y agencias a realizar evaluación. Los objetivos de estos primeros sistemas de SyE de primera generación fueron contribuir a la toma de decisiones presupuestaria, fortalecer la rendición de cuentas dentro del gobierno y asistir a los administradores en los ministerios y agencias. El Departamento de Finanzas diseñó y manejó el sistema de SyE de primera generación. Las evaluaciones fueron obligatorias y debían hacerse cada tres o cinco años para cada programa. Se les requirió a los ministerios sectoriales preparar planes rotativos de tres años para las evaluaciones principales.

Se utilizó una amplia variedad de tipos de evaluación, y para mediados de los años noventa estaban en camino unas 160 evaluaciones en cualquier momento dado. Se exigieron pocos requerimientos formales para recopilar o para presentar información sobre indicadores de desempeño.

Todos los hallazgos de evaluación eran compartidos con el gabinete, que los tomaba muy en serio. De hecho, cerca del 80% de las nuevas propuestas de políticas y dos terceras partes de opciones de ahorro influyeron en la toma de decisiones del gabinete. Otras fortalezas de este sistema fueron el uso intenso de los resultados de las evaluaciones por parte de los departamentos y agencias sectoriales, y el hecho de que la evaluación fuese un esfuerzo de colaboración.

El sistema de primera generación tuvo también debilidades. La calidad de las evaluaciones era dispareja, hubo insuficiencia de apoyo central para capacitación avanzada en evaluación, e insuficientes requerimientos formales para recopilar y presentar información sobre indicadores de desempeño. Los ministerios reclamaron que el sistema era una carga para la administración.

La segunda generación de evaluación en Australia empezó con la elección del nuevo gobierno conservador. Los cambios incluyeron una reducción significativa en el tamaño de la burocracia, el desmantelamiento del sistema de asesoría en políticas para el proceso del presupuesto, la reducción en la supervisión central y las reglas "burocráticas", y la reducción sustancial en el personal del Departamento de Finanzas, que redujo su nómina de asesores en el proceso presupuestario.

Para responder a estos cambios en el gobierno debía cambiar también el sistema de SyE. Se desmanteló la vieja estrategia de evaluación y se "desreguló" la evaluación; es decir, se estimulaba, pero no se exigía. Se enfatizó en el seguimiento del desempeño de productos y efectos directos, que se reportaron al Parlamento tanto *ex ante* como *ex post.*

La Oficina Nacional de Auditoría de Australia informó que en este sistema de segunda generación el desempeño fue altamente inadecuado. La recopilación de datos

fue deficiente debido a los bajos estándares y se utilizaron muy poco los objetivos y las pruebas de comparación. Se recopiló mucha información sobre los productos del gobierno, pero poca sobre los efectos directos y faltó análisis de información sobre el desempeño. Estas deficiencias hicieron que los comités parlamentarios estuviesen muy descontentos con la información que recibían, pero, a pesar de esto, unos cuantos ministerios (familia y servicios comunitarios, educación y salud) aprendieron de las evaluaciones efectuadas con buenas prácticas.

La tercera generación (desde 2006) fue motivada por preocupaciones actuales sobre dificultades para implementar los complejos programas del gobierno, y sobre la "conectividad" (la coordinación entre ministerios y agencias, tanto federales como estatales). Intervino también el deseo del Departamento de Defensa de reconstruir su papel en la asesoría sobre el presupuesto y la política (Mackay, 2007).

Se configuraron dos tipos de revisión para asegurar que el gasto fuera eficiente, eficaz, y de acuerdo con las prioridades del gobierno. Las revisiones estratégicas (siete por año) se centraron en el propósito, el diseño, la gestión, los resultados y las mejoras futuras requeridas en los campos principales de las políticas y el gasto. Las revisiones de programas (también siete anuales) se centraron en programas individuales, y su acuerdo con los objetivos del gobierno, su eficacia, duplicación, superposición y ahorros.

El Departamento de Defensa empezó a manejar las revisiones, y se tomó la decisión de incorporar el sistema como un asunto habitual. El gobierno comprometió US$17 millones para las revisiones en un período de cuatro años. Se contrataron funcionarios públicos retirados para encabezar dos programas piloto. Se mantuvieron los requerimientos de la segunda generación para el marco del control del desempeño.

¿Qué enseñanzas se obtuvieron con la evolución del programa de evaluación de Australia?

- Los asuntos de coordinación, implementación y desempeño (resultados) de los programas constituyen retos permanentes para todos los gobiernos.
- La naturaleza de la toma de decisiones del gobierno determina el nivel de demanda de información (y revisión) de SyE.
- Construir un sistema de revisión de SyE es algo que toma tiempo.
- Es difícil equilibrar las necesidades de información descendentes/centralizadas y ascendentes/descentralizadas.
- La mayor parte de los departamentos no se inclina en forma natural a efectuar evaluaciones, que consideran costosas y peligrosas.

La *Ley sobre Resultados del Desempeño del Gobierno de EUA*. Un desarrollo clave en la evaluación del gobierno en los Estados Unidos, en los últimos 20 años, fue la aprobación del Government Performance Results Act (GPRA) de 1993, con la cual se instituyó la evaluación basada en resultados en todas las agencias del gobierno de EUA. La ley afecta directamente la forma como se realice la evaluación en todas las dependencias del gobierno de EUA. La GPRA es un enfoque que abarca la totalidad del gobierno que se inició con programas piloto antes de efectuar los cambios por fases.

La medición del desempeño en los Estados Unidos empezó con los gobiernos locales en los años setenta, y luego se extendió a los gobiernos estatales y eventualmente al nivel federal, con la promulgación de la GPRA en 1993. El gobierno federal de EUA adoptó la medición del desempeño después de haberse adoptado en los demás niveles del gobierno en los Estados Unidos.

Los objetivos de la GPRA son:

1. Mejorar la confianza de la población estadounidense en la capacidad del gobierno federal, haciendo automáticamente responsables a las agencias federales por el logro de los resultados de los programas.
2. Iniciar la reforma del funcionamiento de los programas con una serie de proyectos piloto en la fijación de objetivos de los programas, midiendo su desempeño frente a esos objetivos e informando al público sobre su progreso.
3. Mejorar la eficacia del programa federal y la rendición de cuentas del sector público, promoviendo un nuevo enfoque en los resultados, la calidad de los servicios y la satisfacción de los usuarios.
4. Ayudar a los administradores federales a mejorar la prestación de los servicios, requiriéndoles que hagan planificación para el cumplimiento de los objetivos, y suministrándoles información sobre los resultados de los programas y la calidad de los servicios.
5. Mejorar la toma de decisiones en el Congreso, suministrando información más objetiva sobre el logro de los objetivos estatutarios y sobre la relativa eficacia y la eficiencia de los programas y gastos federales.
6. Mejorar la administración interna del gobierno federal. La GPRA ordenó a las agencias federales concentrarse en su misión y objetivos, cómo alcanzarlos y cómo mejorar las organizaciones estructurales y los procesos comerciales. Según la ley, se requiere a las agencias remitir planes estratégicos a cinco años para sus programas, y actualizarlos cada tres años. Deben también identificar "factores externos clave" que puedan tener un efecto significativo en la capacidad de la agencia para alcanzar sus metas y objetivos. Las agencias deben publicar informes anuales sobre los resultados de los programas (U.S. Department of Labor, 1993, par. 1).

Las agencias deben también medir sus resultados para garantizar que cumplan sus objetivos, y que las decisiones que tomen se basen en información confiable. Las medidas de su desempeño deben basarse en características relacionadas con el programa y deben ser completas, precisas y coherentes. Deben utilizarse los datos recopilados para mejorar los procesos organizativos, identificar vacíos y fijar objetivos de desempeño (U.S. GAO, 2003).

Según una encuesta de 2003 de 16 programas, a 12 agencias del gobierno de los Estados Unidos, muchos programas federales ya habían utilizado datos de efectos directos recopilados regularmente para el mejoramiento de los programas. Por ejemplo,

se utilizaron datos sobre efectos directos para ejercer acción correctiva, identificar y alentar las "buenas prácticas", motivar y reconocer al personal y planificar y presupuestar.

Al mismo tiempo, en la encuesta se encontraron algunos obstáculos persistentes para usar los datos sobre los efectos directos, entre ellos los siguientes:

- Falta de autoridad o interés para efectuar cambios.
- Uso y conocimiento limitados de los datos sobre los efectos directos.
- Problemas con los datos sobre los efectos directos (obsoletos, sin desagregar, falta de especificidad, necesidad de datos intermedios, etcétera).
- Temor de "perturbar la armonía" (Hatry y otros, 2003).

La GPRA se extendió hasta el año 2003 a fin de integrar el funcionamiento con el presupuesto, y se hicieron también esfuerzos en el gobierno para regular de forma más estrecha la planificación y la presentación de información estratégica y anual de la GPRA. ChannahSorah resume la GPRA como:

> … sencillamente buenos negocios. Sus requerimientos han proporcionado a los departamentos del gobierno herramientas para formas muy básicas de hacer negocios en formas sensibles: fijar objetivos de desempeño y medir los efectos directos a corto y a largo plazos. Cualquier organización que busque ofrecer una mejor calidad de vida, mayor cantidad de servicios y mejor calidad en general de los servicios a los usuarios tiene que tener una visión y una misión, fijar metas y objetivos y medir resultados (2003, pp. 5-6).

En un estudio de 2003, la GAO encontró que muchas agencias tenían problemas significativos para instituir una orientación a los resultados en todas sus dependencias. Los administradores federales encuestados informaron que los dirigentes de la agencia no demostraron de manera consistente un compromiso fuerte para lograr resultados. Además, según estos administradores, las agencias no siempre reconocieron positivamente a los empleados por contribuir a lograr los objetivos estratégicos. La GAO informó también que las organizaciones de alto desempeño buscan desplazar el centro de la administración y la rendición de cuentas, de las actividades y procesos, hacia las contribuciones y el logro de los resultados. Sin embargo, aunque muchos administradores federales encuestados informaron que tenían la responsabilidad por los resultados de sus programas, solamente unos pocos informaron que tenían la autoridad necesaria para que la agencia cumpliese sus objetivos estratégicos. Por último, la GAO encontró que aunque los administradores informaban con mayor frecuencia que tenían medidas de desempeño orientadas a resultados para sus programas, el grado al que estos administradores informaron su uso de la información sobre el desempeño para actividades clave de gestión había disminuido con respecto a encuestas anteriores. El estudio de la GAO destacó la necesidad de transformar culturas organizativas para que fuesen más orientadas a resultados, centradas en el usuario y colaboradoras.

Las principales organizaciones públicas de los Estados Unidos y otros países encontraron que la administración estratégica de capital humano debe ser la pieza central de cualquier iniciativa seria de gestión de cambio, y de los esfuerzos para transformar la

001El camino.indd 50 08/11/2010 03:27:42 p.m.

cultura de las agencias oficiales. Los sistemas de gestión del desempeño son integrales para la administración estratégica del capital humano, y dichos sistemas pueden constituir herramientas clave para maximizar el desempeño, alineando las medidas de desempeño institucional con el desempeño individual, y creando una "línea de visión" entre los objetivos del individuo y los de la organización. Las principales organizaciones utilizan sus sistemas de gestión del desempeño como una herramienta clave para alinear el desempeño institucional con el de cada unidad y cada empleado, obtener resultados, acelerar el cambio, manejar la organización día a día y facilitar la comunicación durante todo el año, de modo que las discusiones sobre el desempeño individual y organizativo se integren y sean continuas.

El enfoque circunscrito

La capacidad de configurar un sistema de evaluación varía mucho según los ministerios. Por eso la estrategia del gobierno total no puede mover simultáneamente a todos ellos, y existe la necesidad de una secuencia para desarrollar estos sistemas en los ministerios. Las innovaciones en un nivel a menudo se filtran horizontal y verticalmente a otros niveles del gobierno.

El *enfoque circunscrito* se concentra en una parte o sector del gobierno, como un ministerio individual. Por ejemplo, México se concentró en el desarrollo social, Jordania en la planeación y Kirguistán en la salud. Trabajar con un ministerio que cuente con un líder fuerte puede constituir el mejor curso de acción en países que carecen de la capacidad de un enfoque de gobierno total.

Enfoque circunscrito: Establecimiento de sistemas de seguimiento y evaluación en una parte o un sector del gobierno a la vez, como en un ministerio individual.

El enfoque mixto

Algunos países, como Irlanda, adoptaron un *enfoque mixto* para la evaluación. Mientras se evalúa de manera integral en algunos campos (como proyectos financiados con fondos estructurales de la UE), otros reciben menos atención. El gobierno de Irlanda empezó creando su sistema de evaluación con un enfoque de enclave, pero se dirigió a enfoque más integral con respecto a programas de gastos del gobierno (Lee, 1999). El enfoque mixto puede constituir también una alternativa válida para algunos países en desarrollo.

Enfoque mixto: Establecimiento integral de sistemas de seguimiento y evaluación en algunas partes del gobierno, mientras otras partes del mismo reciben menos atención.

Aumento para el desarrollo de la capacidad de evaluación en Irlanda. A semejanza de muchos otros países, a finales de los años sesenta, Irlanda se interesó en el análisis racional y su aplicación a la planeación y el presupuesto. Los autores de políticas en el gobierno identificaron la necesidad de estudios objetivos de programas sociales y de desarrollo, y también la necesidad de los participantes de adquirir destrezas para imple-

mentar esos estudios (Friis, 1965; Public Services Organisation Review Group, 1969). Se emprendieron varias iniciativas para desarrollar destrezas en evaluación.

A pesar de estas iniciativas, el alcance de estas evaluaciones fue limitado y tuvieron poca influencia en la toma de decisiones hasta finales de los años ochenta. Esta falta de influencia se debió en parte a la ausencia de una tradición fuerte de evaluación de políticas y programas en Irlanda y parcialmente a que las evaluaciones se realizaron en un tiempo de crisis económica, cuando la evaluación como herramienta de buena gobernanza no se consideraba tan importante como la necesidad de controlar el gasto público.

Una excepción a esa falta de influencia fue la del gasto de la UE en Irlanda. Los fondos de la UE se aplican a través de varios programas operativos, que se ejecutan en un plan conjunto dentro del Irish-EU Community Support Framework (CSF, Marco de Apoyo Irlanda-Comunidad de la UE). La UE, fuente importante de apoyo financiero, demandó que la evaluación fuera consistente y sistemática. Así, las evaluaciones de programas financiados por ella afectaron significativamente dos campos principales de política: a) la capacitación industrial y los planes de creación de empleo, y b) programas contra la pobreza y otros programas de desarrollo comunitario. El campo del mercado laboral tendió a enfocarse en la medición cuantitativa de resultados, mientras las iniciativas de desarrollo comunitario se centraron en métodos cualitativos referentes a la descripción y no a la medición de resultados (Boyle, 2005).

Entre 1989 y 1993, se fundaron dos unidades de evaluación independiente; una la fundó el European Social Fund (Fondo Social Europeo) y la otra una unidad de evaluación industrial. Desde 1989, la evaluación de los Fondos Estructurales de la UE constituyó un requisito formal para quienes recibiesen asistencia que condujera a desarrollos ulteriores en evaluación. Entre 1994 y 1999 se estableció una unidad de evaluación central en el Departamento de Finanzas; una tercera unidad de evaluación se estableció con el fin de cubrir las evaluaciones en agricultura y desarrollo rural y se nombraron evaluadores externos para los gastos de programas operativos y el plan CSF. Entre 1999 y 2006 se renovó el interés en la evaluación del gasto público en Irlanda, se incrementó la capacidad de la unidad de evaluación central para asumir responsabilidades adicionales y se eliminaron las unidades de evaluación independiente. Se contrató a evaluadores externos para la evaluación de mitad de período de los programas operativos y el plan nacional de desarrollo (Boyle, 2005).

Adopción de un nuevo enfoque en la evaluación en Francia. Hasta 2001, Francia pertenecía al grupo de países de la OCDE que más tardó en adoptar un sistema maduro de evaluación. En realidad, Francia se rezagó con respecto a muchas economías de transición y países en desarrollo en este aspecto. Se hicieron varios intentos de reformas incrementales durante los últimos años de las décadas de 1980 y de 1990.

Entonces, en 2001, el gobierno aprobó una legislación arrolladora, que reemplazó la legislación constitucional financiera de 1959, eliminó el presupuesto por renglones e instituyó un nuevo enfoque para los programas. El nuevo estatuto constitucional, con

001El camino.indd 52 08/11/2010 03:27:42 p.m.

varias fases en un período de cinco años (2001-06), tenía dos objetivos primordiales: a) reformar el marco de la administración pública, para convertirlo en un marco orientado a resultados y desempeño; y b) fortalecer la supervisión legislativa. Como el entonces primer ministro Lionel Jospin anotó, "la presentación del presupuesto en forma de programas que agrupen el gasto según políticas públicas mayores debería darles tanto a los miembros del Parlamento como a los ciudadanos un cuadro claro de las prioridades del gobierno y del costo y los resultados de su acción" (República de Francia, 2001). Se identificaron unos 100 programas y se presupuestaron recursos financieros para ellos. Cada presupuesto de programa sometido a la legislatura debía tener una declaración de objetivos precisos y de indicadores de desempeño. Los administradores públicos tuvieron mayor libertad y autonomía en cuanto a la asignación de los recursos, pero a cambio se responsabilizaron más por los resultados. Así, el nuevo proceso de presupuesto se orientó a los resultados.

Las solicitudes de presupuesto de fondos adicionales debían incluir planes de desempeño anual, detallando los resultados esperados frente a los reales para cada programa, y también se incluyeron los informes sobre desempeño anual en las revisiones del presupuesto. Estos pasos se dieron para mejorar la capacidad del legislador de evaluar el desempeño de los programas del gobierno.

Esta iniciativa de reforma alteró algunas relaciones políticas e institucionales dentro del gobierno francés, otorgándole a la legislatura mayores poderes presupuestarios. "El artículo 40 de la Constitución prohibía anteriormente a los miembros [de la legislatura] poner sobre la mesa enmiendas que aumentaran el gasto y redujeran las rentas. Pueden cambiar la distribución de las apropiaciones entre los programas en una misión dada". La legislatura puede votar sobre estimados de rentas, apropiaciones para misiones individuales, límites al número de creación de empleos estatales y cuentas especiales y presupuestos específicos. Además, los comités legislativos de finanzas tienen responsabilidades de control y supervisión referentes al presupuesto.

Los funcionarios públicos reaccionaron inmediatamente ante los cambios. Había una nueva burocracia de control, unos nuevos contadores, más auditorías, más cuestionarios sobre las oficinas de auditoría y los inspectores, más solicitudes de presentación de información, etcétera. Los administradores tuvieron dificultad para adaptarse a las restricciones de obtener resultados de producto (cantidad) mientras se ignoraba la calidad de los servicios, que no aparecía en los objetivos.

En cuanto a la calidad de los servicios, "ningún mecanismo de competencia o de presiones consumistas fuertes hacen posible garantizarla" (Trosa, 2008, p. 8). Las sociedades son muy complejas. Algunas personas necesitan asistencia financiera, otras necesitan confianza y asumir responsabilidad; pero otras viven felizmente. La cuestión es cómo resumir estas tensiones en una fórmula (Trosa, 2008).

Combinar el modelo anterior de evaluación con el nuevo no permitía libertad de administración, creatividad o innovación. Trosa indica que se precisa un modelo alternativo. Otra enseñanza de la experiencia francesa es que "no es posible mejorar la administración interna sin vincularla a la gobernanza interna de los sectores públicos"

001El camino.indd 53 08/11/2010 03:27:42 p.m.

(Trosa, 2008, p. 2). Según Trosa, no es necesario demoler el nuevo sistema, sino más bien ampliarlo, discutiendo claramente los propósitos requeridos, y estimulando al mismo tiempo la lógica de la acción, y no simplemente el uso de herramientas.

La evaluación en los países en desarrollo

Los países en desarrollo se enfrentan a problemas semejantes, y también diferentes a los que enfrentan los países desarrollados cuando empiezan a adoptar y construir sus propios sistemas de evaluación. Para establecer un sistema de evaluación y que surta efectos en cualquier país, se necesita que haya partes interesadas y compromiso con la transparencia y la buena gobernanza. La demanda y propiedad de un sistema de evaluación pueden ser más difíciles de establecer en los países en desarrollo.

Una débil voluntad política y una capacidad institucional retrasan el progreso. Las dificultades en la cooperación y coordinación entre los ministerios pueden impedir que progrese la planeación estratégica. En realidad, la falta de suficiente cooperación y coordinación gubernamental es un factor para esto tanto en los países desarrollados como en los en desarrollo.

Para que puedan surgir y madurar los sistemas de evaluación, es necesario que haya voluntad política en el gobierno, y que los líderes en altos cargos estén dispuestos a asumir los riesgos políticos de defender la evaluación. La presencia de un líder o de líderes nacionales puede hacer mucho para que un país desarrolle y sostenga un sistema de evaluación. A la inversa, no se conoce ningún caso en el que haya surgido un sistema de SyE en el sector público de un país en desarrollo, sin la presencia de un líder.

Muchos países en desarrollo pugnan aún por estructurar instituciones eficaces fuertes. Algunos precisan una reforma en el servicio civil o del marco regulador y legal. Con este fin, la comunidad de desarrollo internacional intenta mejorar los cimientos para apoyarlos. El problema está en formar instituciones, emprender reformas administrativas y el servicio civil y/o rediseñar el código legal y regulador, estableciendo al mismo tiempo sistemas de evaluación. Instituir sistemas de evaluación puede contribuir a informar y orientar al gobierno para iniciar reformas necesarias en todos estos campos.

Los países en desarrollo deben primero tener o establecer un fundamento para la evaluación, y muchos se están moviendo en esta dirección. Establecer un fundamento significa contar con sistemas y datos estadísticos básicos y también sistemas presupuestarios clave. Los datos y la información deben ser de calidad y cantidad apropiadas. Al igual que los países desarrollados, los países en desarrollo deben conocer sus condiciones básicas: dónde se encuentran actualmente en relación con un programa o una política dados. Se necesita una capacidad en la fuerza laboral para desarrollar, apoyar y sostener estos sistemas. Los funcionarios deben estar capacitados en la recopilación moderna de datos, métodos de control y análisis, y éstos pueden ser problemas difíciles en muchos países en desarrollo (Schacter, 2002).

001El camino.indd 54 08/11/2010 03:27:42 p.m.

Como respuesta a estos problemas, muchas organizaciones de ayuda han incrementado sus esfuerzos para formar capacidad institucional. Los métodos incluyen asistencia técnica y financiera para construir sistemas estadísticos, capacitación en seguimiento y evaluación, evaluaciones de diagnóstico y resultados y sistemas de presupuesto basados en el desempeño. La tendencia hacia las Estrategias de Asistencia al País basadas en resultados contribuye a modelar las prácticas. La ayuda a los países en desarrollo en la producción de estrategias para reducir la pobreza dirigidas por el mismo país contribuye también a formar esa capacidad.

Como parte de los esfuerzos para apoyar la capacidad local en los países en desarrollo, las organizaciones de desarrollo están dando los pasos para crear redes, tales como redes de computadores en línea y comunidades participativas que compartan pericias e información. Ejemplos de esto son la Development Gateway y la Asian Community of Practice. Se puede argumentar que las circunstancias de un país particular son únicas, y que la experiencia de un país no se trasladará necesariamente a otro, pero una vez aceptado que hay poco conocimiento genérico sobre el desarrollo, que todo conocimiento debe acumularse y luego analizarse, modificarse, desensamblarse y recombinarse para ajustarlo a las necesidades locales, la fuente es intrascendente. El nuevo lema es "explore globalmente y reinvente localmente" (Fukuda-Parr, Lopes y Malik, 2002, p. 18).

El Programa Internacional para la Capacitación en Evaluación para el Desarrollo

En 1999, el Independent Evaluation Group del Banco Mundial realizó una encuesta para identificar la capacitación en evaluación para el desarrollo, y encontró que era escasa, con excepción de unos pocos programas especiales para organizaciones de desarrollo. Estos hallazgos llevaron a que se creara el Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET) en 2001, como parte de un esfuerzo mayor para formar capacidad en evaluación en países en desarrollo y en organizaciones especializadas en temas de desarrollo.

El IPDET capacita profesionales que trabajan, o están por trabajar, en el diseño y la implementación de evaluaciones en el contexto para el desarrollo. El programa de capacitación, que se efectúa cada año en Ottawa, Canadá, en el campus de la Carleton University, y cuenta con apoyo de otras organizaciones, dura cuatro semanas y es una colaboración del Independent Evaluation Group del Banco Mundial, la Faculty of Public Affairs de Carleton University y otras organizaciones, entre ellas la Canadian International Development Agency (CIDA), el Department for International Development (DfID) del Reino Unido, la Agencia Suiza de Cooperación para el Desarrollo (SDC), la Agencia noruega de Cooperación para el Desarrollo (NORAD), el Centro Internacional de Investigaciones para el Desarrollo (CIID), el Geneva International Centre for Humanitarian Demining (GICHD), la Agencia Sueca de Cooperación para el Desarrollo Internacional (SIDA), la Commonwealth

001El camino.indd 55 08/11/2010 03:27:42 p.m.

Secretariat, el Dutch Ministry of Foreign Affairs, el Banco Africano de Desarrollo (BAfD) y la Danish International Development Agency (DANIDA).

El curso del IPDET se inicia con un programa central de dos semanas que dedica especial atención al control y la evaluación de la implementación de estrategias de reducción de la pobreza, y hace hincapié en el SyE basado en resultados y en la participación de los interesados. El programa central ofrece más de 80 horas de instrucción con herramientas, casos de estudio, grupos de discusión y lecturas. Casi la tercera parte del tiempo de instrucción se dedica a sesiones de grupos de trabajo estructurados, que ofrecen a los participantes con intereses similares la oportunidad de trabajar juntos sobre asuntos de evaluación para el desarrollo del mundo real, y producen un diseño de evaluación preliminar para un programa que uno de los participantes en el grupo debe evaluar a su regreso al trabajo.

Después del programa central de dos semanas se ofrecen otras dos semanas de capacitación adaptadas a las necesidades individuales, y que se realizan por medio de 30 talleres que dirige un grupo de instructores muy respetados y bien conocidos. Algunos ejemplos de estos talleres son:

- Diseño de evaluaciones de impacto bajo restricciones.
- Diseño y construcción de sistemas de SyE basado en resultados.
- Comprensión de los enfoques del Banco Mundial para la evaluación de países, sectores y proyectos.
- Uso de combinaciones de métodos para evaluaciones de desarrollo.
- Uso de SyE participativo.
- Evaluación de situaciones de posconflicto y realización de evaluaciones internacionales conjuntas.
- Evaluación de programas de VIH/Sida.
- Evaluación de poblaciones ocultas y marginadas.

El IPDET es uno de los pocos programas sobre evaluación para el desarrollo que se ofrece anualmente, y ha capacitado a más de 2.000 profesionales provenientes de ministerios del gobierno, organizaciones de ayuda bilateral y multilateral, ONG y otras entidades de desarrollo.

El programa central del IPDET se ofrece ya a niveles regional y nacional. Se han efectuado versiones adaptadas de éste, por ejemplo en Australia (para la comunidad australiana de ONG), Botsuana, Canadá, China, India, Sudáfrica, Túnez, Tailandia, Trinidad y Tobago, y Uganda. Estos programas regionales, de menor duración y conocidos como "mini-IPDET", son en gran parte interactivos y en ellos se utilizan presentaciones, grupos de discusión, ejercicios grupales y casos de estudio. Los programas están diseñados para maximizar las oportunidades de aprendizaje, a la par con un enfoque en discusiones de problemas del mundo real y aprendizaje de soluciones prácticas a esos problemas.

001El camino.indd 56 08/11/2010 03:27:42 p.m.

Basado en el éxito demostrado del IPDET mediante evaluaciones anuales y de impacto, el Independent Evaluation Group entró en asociación con el Ministerio de Finanzas de China (Departamento Internacional), el Asia-Pacific Finance Department Center y el Asian Development Bank sobre la primera oferta regional institucionalizada. El programa, denominado Shanghai IPDET (Shipdet), se ofrece dos veces por año, una de nivel nacional y otra de nivel regional. Su creación ilustra los esfuerzos para ayudar a desarrollar capacidad en evaluación dentro de países y regiones en desarrollo.

Nuevos sistemas de evaluación

Los intentos de desarrollar un sistema de evaluación e ilustrar sobre la asignación de los recursos y los resultados reales pueden encontrar resistencia política, hostilidad y oposición. Dada la naturaleza de muchos gobiernos de países en desarrollo, la formación de un sistema de evaluación puede también llevar a reconfigurar las relaciones políticas. Para crear un sistema de evaluación maduro se requieren interdependencia, alineamiento y coordinación entre los diversos niveles de gobierno, y lograr tales condiciones representa un gran reto. En muchos países en desarrollo, los gobiernos apenas están interconectados y todavía en la labor de formar culturas administrativas sólidas y sistemas financieros transparentes. Como resultado, algunos gobiernos solamente cuentan con una vaga información en cuanto al monto y la asignación de los recursos disponibles, y necesitan más información sobre si estos recursos se están utilizando para los propósitos previstos. En tales entornos, medir el desempeño del gobierno es únicamente un ejercicio de aproximación.

Muchos países en desarrollo y desarrollados están todavía trabajando para vincular el desempeño con el marco o la estrategia de los gastos públicos. De no formarse esos vínculos, no existe forma de determinar si las asignaciones presupuestarias que los programas de apoyo están ejecutando en último término tendrán éxito.

Algunos países en desarrollo están empezando a progresar en este campo. Por ejemplo, en los años noventa Indonesia comenzó a vincular la evaluación con su proceso de asignación presupuestaria anual. "La evaluación se considera como una herramienta para corregir políticas y programas de gasto público mediante enlaces más directos al Plan Nacional de Desarrollo y al proceso de asignación de recursos" (Guerrero, 1999, p. 5). Algunos países de medianos ingresos, entre ellos Brasil, Chile y Turquía, han progresado en la vinculación de los gastos con los objetivos de productos y efectos directos. El gobierno de Brasil emite informes oficiales separados sobre los objetivos de los efectos directos (OCDE y PUMA, 2002).

Muchos países en desarrollo operan todavía con dos sistemas presupuestarios, uno para los gastos recurrentes y otro para los gastos de inversión/capital. Hasta hace poco tiempo, el Ministerio de Finanzas de Egipto supervisaba el presupuesto de gastos y su Ministerio de Planeación supervisaba el presupuesto de capital. La consolidación de estos presupuestos dentro de un solo ministerio le ha facilitado al gobierno considerar

001El camino.indd 57 08/11/2010 03:27:42 p.m.

un sistema de evaluación de base amplia para garantizar que se cumplan las metas y objetivos del país.

Dadas las dificultades particulares de establecer sistemas de evaluación en los países en desarrollo, puede ser preferible adoptar un enfoque de enclave, o parcial (en el cual unos pocos ministerios o departamentos realizan programas piloto y adoptan sistemas de evaluación) al enfoque que abarca la totalidad del gobierno. Intentar establecer un enfoque de gobierno total para la evaluación podría ser muy ambicioso. Por ejemplo, una evaluación de la preparación efectuada por el Banco Mundial para Kirguistán, recomendó que se apoyara al Ministerio de Salud (que ya contaba con alguna capacidad de evaluación) como modelo potencial para la eventual implementación en todo el gobierno de un sistema de evaluación.

China, Malasia y Uganda han preferido el enfoque circunscrito, y más adelante se describen sus esfuerzos.

Evaluación motivada por el crecimiento en China. La evaluación es un fenómeno relativamente nuevo en China, hasta tal punto que antes de principios de los años ochenta era casi desconocido. Este desconocimiento de la evaluación era un reflejo de la orientación de las ciencias sociales, en ese tiempo, de la virtual ausencia de literatura publicada sobre evaluación en chino, y de la falta de contactos sistemáticos de los ciudadanos chinos con los profesionales de la evaluación en otras partes del mundo.

No obstante, los chinos efectuaron algunas actividades relacionadas con la evaluación, tales como análisis político, estudios económicos y de administración, investigación mediante encuestas, revisiones sobre finalización de proyectos y síntesis de experiencias. En algunos institutos de política e investigación económica existía capacidad institucional, técnica y analítica en ciencias sociales.

Sin embargo, sólo hasta 1992 las agencias centrales principales, como la Administración Estatal de Auditoría, el Ministerio de Finanzas y la Comisión Estatal de Planeación, empezaron a desarrollar y presentar propuestas específicas para formar capacidad en SyE del desempeño en el Consejo de Estado. El Centro para la Evaluación de Ciencia y Tecnología y el Departamento de Evaluación de Política y Operaciones de los Países Bajos realizaron una primera evaluación conjunta de programas de ciencia y tecnología (NCSTE e IOB, 2004).

Con el capital y la ayuda para el desarrollo de China en los últimos 20 años, el país ha experimentado un incremento en su capacidad y conocimiento en análisis tecnológico y de ingeniería, análisis financiero, análisis y modelado económico, análisis de impacto ambiental, análisis de sostenibilidad y estudios de implementación. En realidad es sorprendente la rapidez con que ha surgido la capacidad en China.

En este país, la fuerza motriz para la evaluación son el surgimiento masivo y sostenido del desarrollo nacional y el crecimiento económico. El producto interno bruto (PIB) aumentó más de un 7,8% por año durante los nueve años anteriores a 2008. El interés en los asuntos de evaluación proviene de la preocupación de China por el desarrollo. Algunas agencias centrales, incluyendo la Compañía Internacional de Consultoría en

Ingeniería de China, una firma de consultoría de propiedad del Estado, el Ministerio de Construcción y el Banco de Desarrollo del Estado, han formado capacidad de evaluación a los más altos niveles de esas organizaciones.

Aunque la mayor parte de la evaluación es evaluación *ex post* de proyectos, cada vez se reconoce más que los temas de evaluación deben incorporarse en todas las etapas del ciclo de los proyectos de desarrollo, y cada vez hay mayor conciencia dentro de China de que la función de evaluación es aplicable a todas las etapas del ciclo de los proyectos. Ya hay interés en vincular la evaluación a la formulación y la implementación de proyectos y programas, y existen evaluaciones en curso, aunque son todavía infrecuentes.

Un ejemplo de evaluación fue la que se efectuó en 2006, cuando China incorporó por primera vez un componente sistemático de SyE en su plan quinquenal. Este sistema, incluido en el Undécimo Plan Quinquenal, se basa en los 10 pasos que se identifican en Kusek, Rist y White (2004).

En abril de 2006, China lanzó semestralmente el Shipdet con el fin de capacitar evaluadores nacionales y regionales. Los asociados en el programa de capacitación son el Ministerio de Finanzas, el Banco Mundial, el Banco Asiático para el Desarrollo y el Asia-Pacific Finance and Development Center. China está, además, fijando los fundamentos para la evaluación, aunque el edificio aún no está terminado.

En la estructura de gobierno de China y su jerarquía administrativa, deben llevarse a cabo varias tareas básicas para que la evaluación continúe su desarrollo, así:

- Establecer una fuerte organización central para la administración y la coordinación general de la evaluación.
- Establecer unidades de evaluación formal, políticas y lineamientos en todos los ministerios y bancos pertinentes.
- Reconocer que el momento es adecuado para que los gobiernos provinciales y locales inicien sus propias evaluaciones.
- Establecer en la Administración Estatal de Auditoría un proceso de auditoría de la función de evaluación, de modo que puedan efectuarse la supervisión y la auditoría continua de las evaluaciones emprendidas dentro de los ministerios de línea y las políticas y lineamientos sobre evaluación, emitidas por las organizaciones de evaluación central, y los ministerios, provincias y bancos pertinentes.
- Desarrollo de métodos avanzados de evaluación en las unidades y entidades organizativas.
- Fortalecimiento de la función de seguimiento y evaluación en las agencias de inversión.
- Desarrollo de una oferta de evaluadores bien capacitados para los muchos ministerios, provincias y bancos nacionales que están por iniciarse en el campo de la evaluación (Houqi y Rist, 2002).

China ha identificado la importancia de la demanda de resultados en las evaluaciones. Este tema clave representa un reto para muchos países que no tienen una tradición de gobierno transparente.

Presupuesto basado en resultados, formación de la nación y competitividad global en Malasia. Entre los países en desarrollo, Malasia ha estado a la vanguardia en cuanto a reformas de la administración pública, en especial en los campos del presupuesto y las finanzas. Las reformas se iniciaron en los años sesenta como parte de un esfuerzo del gobierno para desarrollar el país de manera estratégica. En vista de que se consideraba al sector público como el vehículo principal para el desarrollo, se dio especial importancia a fortalecer el servicio civil mediante reformas administrativas.

En 1969, Malasia adoptó el Sistema de Planeación, Programación y Presupuesto (PPBS), que continuaba usándose en los años noventa. Ese sistema reemplazó al de presupuesto por renglones con el presupuesto basado en resultados. Aunque las agencias utilizaban la estructura de programas y actividades, en la práctica la ejecución todavía era semejante al presupuesto por renglones y un enfoque incremental. La reforma presupuestaria se centró en aumentar la rendición de cuentas y la disciplina financiera en las agencias del gobierno, encargadas de realizar los planes de desarrollo socioeconómico de Malasia. Además, el gobierno inició varias reformas adicionales, incluyendo esfuerzos para mejorar el cumplimiento financiero, la administración de calidad, la productividad, la eficiencia en las operaciones del gobierno y la gestión de los esfuerzos para el desarrollo nacional.

Los esfuerzos de reforma presupuestaria de Malasia se vincularon junto con los de formación de la nación y de competitividad global. Una de las fuerzas impulsoras de estos esfuerzos de reforma fue la vinculación de la reforma al programa Visión 2020, que aspira a hacer de Malasia un país totalmente desarrollado para el año 2020.

En 1990, el gobierno reemplazó el PPBS con el Modified Budgeting System (MBS). Con el PPBS los vínculos entre productos e insumos eran mínimos, y las políticas continuaban financiándose aun cuando no se medían los resultados de manera sistemática. Con el MBS, se dio mayor importancia a los productos y al impacto de los programas y actividades del gobierno.

A finales de los años noventa, Malasia desarrolló su sistema integrado de administración basado en resultados (IRBM, por sus iniciales en inglés). Los componentes del sistema fueron un sistema de presupuesto basado en resultados y un sistema de evaluación de desempeño del personal también basado en resultados. Además, Malasia desarrolló otros dos sistemas basados en resultados: un sistema de información para la administración (MIS, por sus iniciales en inglés) y un marco de SyE, para complementar el sistema de administración (Thomas, 2007).

El sistema de IRBM proporciona un marco para la planeación, ejecución, el control y la presentación de información sobre el desempeño organizativo. También tiene capacidad para vincular el desempeño organizativo con el desempeño del personal. Varios países (entre los que figuran Afganistán, Botsuana, India, Mauricio y Namibia) están

integrando sistemas de IBRM por etapas, con presupuesto y SyE basados en resultados en primer plano. En Malasia, el sistema presupuestario es el motor principal del sistema de IRBM (Thomas, 2007).

El MIS y el sistema de SyE, que aportan la dimensión de medición del desempeño al marco de planeación estratégica, hacen más dinámico al IRB. El MIS y SyE están estrechamente vinculados para garantizar que el sistema produzca la información correcta para las personas correctas en el momento correcto. Los indicadores deben ser tanto operativos como basados en resultados. En Malasia se desarrolló y utilizó una versión electrónica del marco de la administración integrada del desempeño (Thomas, 2007).

El gobierno malayo identificó varias enseñanzas aprendidas de esta experiencia:

- Un programa de formación de capacidad para niveles clave del gobierno debe ser sostenido.
- El control y la presentación de información consumen mucho tiempo.
- El proceso de planeación del desempeño debe fortalecerse de manera que sea más integral que incremental.
- Las recompensas y las sanciones no se proporcionan en todos los niveles.
- La integración con otras iniciativas es limitada (Rasappan, 2007).

Propuso también estas recomendaciones:

- Trabajar hacia vínculos verticales y horizontales totales.
- Evitar la desilusión en los niveles de política y operativo.
- Modificar y fortalecer las políticas y sistemas de apoyo.
- Trabajar por un sistema integrado de gestión basada en resultados que se concentre en el desempeño del gobierno total (Rasappan, 2007).

Pese a que Malasia ha estado a la vanguardia de la administración pública y las reformas presupuestarias, sus reformas no han sido suaves ni consistentes a través del tiempo. No obstante, el MBS constituyó una iniciativa importante por parte del gobierno, que demostró previsión, capacidad de innovación, dinamismo y compromiso para garantizar el valor por el dinero en los proyectos y políticas implementados (Banco Mundial, 2001).

La reducción de la pobreza como impulso para la evaluación en Uganda. El gobierno de Uganda se comprometió con una prestación eficaz de servicios públicos como soporte de sus prioridades de reducir la pobreza. El reconocimiento de la eficacia en la prestación de los servicios como un imperativo de la administración nacional para el desarrollo constituye una fuerte evidencia de su compromiso con los resultados. Este compromiso también se evidencia en varias prioridades y actividades en curso de la administración pública.

En la última década, Uganda experimentó una reforma económica integral, y logró la estabilización macroeconómica. Como respuesta al Comprehensive Development Framework desarrolló el Plan de Acción para la Erradicación de la Pobreza (PEAP), que

se encuentra incorporado a su Documento de Estrategia de Reducción de la Pobreza. El PEAP convoca a reducir la tasa de pobreza absoluta, de un 44% en que se encontraba a finales de los años noventa, al 10% para 2017. El PEAP y los ODM son en gran parte semejantes en cuanto a su enfoque, y comparten el objetivo general de responsabilizar al gobierno y sus asociados en el desarrollo del progreso de éste.

Uganda fue el primer país en ser declarado elegible y beneficiado de la Iniciativa para los Países pobres altamente endeudados (PPAE). El año 2000 calificó para incremento en la ayuda de la PPAE, como reconocimiento a la eficacia de su estrategia de reducción de la pobreza, su proceso consultivo que incluyó a la sociedad civil, y su compromiso continuo con la estabilidad macroeconómica.

Uganda introdujo nuevas medidas para que el proceso presupuestario fuese más abierto y transparente a las partes interesadas internas y externas. El gobierno está modernizando sus sistemas fiscales, e iniciando un programa de descentralización de la planeación, la administración de los recursos y la prestación de servicios a las localidades. El Ministerio de Finanzas, Planeación y Desarrollo Económico (MFPED) está introduciendo también el presupuesto orientado a productos. Además, se fortalecerán las instituciones del gobierno, que serán más responsables ante el público.

El país experimenta aún dificultades en la coordinación y armonización con respecto a la evaluación y el PEAP. "La característica más obvia del sistema PEAP SyE es la separación del control de la pobreza y el control de los recursos, si bien ambos son coordinados por el MFPED. Las dos ramas del SyE tienen actores e informes separados y utilizan criterios diferentes de evaluación. El control de los recursos financieros se asocia con insumos, actividades y, cada vez más, con productos; mientras el control de la pobreza se basa en el análisis de los efectos directos generales sobre la pobreza" (Hauge, 2001). Otros problemas de coordinación de evaluación se refieren a la creación de una nueva National Planning Authority (Autoridad Nacional de Planeación) y a los grupos de trabajo sectoriales.

A finales de 2007, el Despacho del Primer Ministro (OPM), presentó un documento de trabajo para la discusión de SyE del plan nacional de desarrollo. Dos de los objetivos de este documento fueron los de revisar las fortalezas y debilidades del PEAP, y proponer un camino de avance para SyE del nuevo plan nacional (Uganda OPM, 2007a).

El documento de trabajo informó sobre estos problemas con el sistema:

- No están claros los "efectos directos y productos de los ministerios sectoriales, indicadores mensurables con bases y objetivos asociados, sistemas eficientes de control y uso estratégico de la evaluación para determinar el desempeño y la causalidad" (Uganda OPM, 2007b, p. 4).
- La rendición de cuentas se basa en el gasto y no en medidas sustantivas de desempeño.
- No hay equilibrio entre la cantidad de datos recopilados, la demanda de datos y la capacidad de utilizarlos.

- Como resultado de la duplicación y descoordinación del control se ha creado una carga compleja y formidable de actividad de inspección, recopilación de datos para indicadores y formatos de informes. El resultado de esto es un gran volumen de datos en cumplimiento con las reglas y regulaciones que no proporcionan una base clara para evaluar el valor por el dinero y la efectividad en costos en el desempeño del sector público. Según un estudio realizado por la Uganda OPM (2007a, p. 6) "las razones de la deficiente coordinación y duplicación de esfuerzos se relacionan más con la estructura de incentivos del servicio civil, donde las actividades de control las motiva en parte el deseo de pagos por día como complemento importante del salario".
- La falta de incentivos y los problemas de superposición en los mandatos sobre temas de planeación y SyE relacionados han dificultado convocar un grupo de trabajo nacional de SyE que aborde los problemas de SyE.
- Aunque se efectúan numerosas evaluaciones, usualmente se llevan a cabo dentro de sectores y ministerios sin utilizar los estándares comunes.
- A nivel local, las personas todavía no se consideran participantes en la toma de decisiones.

En el documento de trabajo se propusieron igualmente estas recomendaciones:

- Hacer depender las asignaciones presupuestarias del logro de resultados.
- Analizar el establecimiento de acuerdos sobre servicios públicos o contratos basados en el desempeño.
- Suministrar información sobre los resultados en forma oportuna y utilizable para los autores de políticas.
- Asegurar que la demanda de información y de datos se refleje en su suministro.
- Establecer mecanismos para el control y garantía de la calidad.
- Garantizar que el análisis sea útil para los autores de políticas.
- Separar las funciones de seguimiento y evaluación.
- Clarificar los papeles y responsabilidades en el gobierno para planeación, control, evaluación y otras funciones relacionadas con la garantía de la calidad.

Con respecto a la evaluación futura, Uganda se enfrenta al reto de mantener el seguimiento y el aprendizaje de su progreso hacia la reducción de la pobreza, por medio del PEAP y la Estrategia Nacional de Reducción de la Pobreza. La evaluación no puede estar aislada de las prácticas de la toma de decisiones y de los incentivos que soportan los sistemas y procesos para el desarrollo nacional (Hauge, 2001).

Implicaciones de los problemas emergentes para el desarrollo

Los problemas emergentes para el desarrollo hacen de la evaluación algo más complejo. En esta sección se ofrece una visión general de estos problemas, destacando las implicaciones para la evaluación.

Patton (2006) inició una discusión de las tendencias recientes en evaluación, identificándola como un bien público global y describió el crecimiento de organizaciones profesionales, asociaciones y sociedades de evaluación alrededor del mundo, y los estándares y lineamientos que establecen estas organizaciones. Además anotó el desarrollo de más de 100 modelos nuevos para la evaluación como una tendencia emergente.

Patton utilizó una analogía para ilustrar la emergente complejidad de la evaluación. En el pasado, los evaluadores podían con mucha frecuencia seguir una especie de receta al efectuar las evaluaciones, y él describe los méritos de la receta como sigue:

- Las recetas se ensayan, para asegurar su posibilidad de replicación.
- Aunque no se necesita ninguna pericia, saber cocinar aumenta la posibilidad de éxito.
- Las recetas producen productos estándar.

Las recetas funcionan bien en la cocina, pero no producen resultados estándar en el desarrollo, cuando se requiere un modelo en el cual el evaluador debe reaccionar a cuestiones complejas. La analogía de Patton para la tendencia emergente en la evaluación para el desarrollo es la de criar un hijo. Una receta es un proceso que se sigue paso a paso. En contraste, criar un hijo es un proceso muy complejo, en donde los custodios utilizan el conocimiento para ayudarles a tomar decisiones y reaccionar a nuevas situaciones.

Patton describe otra tendencia en la evaluación para el desarrollo: la de pasar a situaciones más formativas. La clase de situaciones formativas que son evaluaciones en las cuales:

- Se especifican los efectos directos previstos y esperados, pero las técnicas de medición se prueban en programas piloto.
- Se hace hipótesis de un modelo para obtener efectos directos, se somete a prueba y se refina.
- La ejecución de la intervención no es estandarizada, sino que se estudia y mejora cuando se resuelven los problemas de la intervención (una especie de enfoque iterativo).
- La atribución se formula con la posibilidad de probar la causalidad como parte del problema.

Otra tendencia reciente es la de ir más allá de los estudios de evaluaciones individuales, a las corrientes (Rist y Stame, 2006). Rist y Stame describieron cómo la comunidad de evaluación depende ya de sistemas de conocimiento evaluativos, no de evaluadores o evaluaciones individuales, para producir conocimiento evaluativo. Un aspecto básico de su tesis es que simplemente acumular más y más informes de evaluaciones no causa impacto, o causa muy poco, en el conocimiento resultante que podría obtenerse con una síntesis de estos mismos estudios.

Lo que sucede en el desarrollo afecta la evaluación. La agenda para el desarrollo continuará evolucionando como respuesta a los problemas actuales y los emergentes,

incluyendo la globalización, la creciente incidencia de conflicto alrededor del mundo, el terrorismo y el lavado de dinero, la brecha creciente entre los ricos y pobres en el mundo, el creciente número de participantes en la escena para el desarrollo, el impulso hacia la reducción de la deuda y el nuevo enfoque en una mejor gobernanza. El tratamiento de estos problemas plantea nuevas demandas al evaluador.

El impulso global hacia un desarrollo integral, coordinado y participativo y la demostración de resultados tangibles presenta también nuevos desafíos a la comunidad de evaluación para el desarrollo. Ha habido desplazamientos significativos del desarrollo parcial al integral, del enfoque individual al coordinado (asociaciones), de la promoción del crecimiento a la reducción de la pobreza y del enfoque en la ejecución al enfoque en los resultados. Con respecto al desarrollo integral, por ejemplo, las entidades donantes bilaterales y multilaterales "deben ya posicionar cada proyecto dentro de un contexto más amplio y examinar su sostenibilidad y efectos potenciales en la sociedad, la política y la economía general" (Takamasa y Masanori, 2003, p. 6). Como anotan ellos:

> Los teóricos para el desarrollo llegan también a creer que el factor más importante para el desarrollo económico no es el capital sino las políticas y las instituciones apropiadas. La causa de este cambio fue el tremendo impacto que economistas como North…, Stiglitz… y Sen… produjeron en la disciplina de la economía, incluida la economía para el desarrollo. Estos desarrollos dieron como resultado la situación actual en la que el tema central de la ayuda internacional para el desarrollo es la reducción de la pobreza en un sentido amplio, que incluye la expansión de la dignidad humana y la libertad política y económica para los pueblos de los países en desarrollo.

Los ODM constituyen una manifestación concreta de esta nueva corriente de pensamiento sobre el desarrollo. El *Informe sobre el Desarrollo Mundial 2005* (Banco Mundial, 2005b) se concentra en lo que pueden hacer los gobiernos para crear mejores climas de inversión en su sociedad, midiendo el progreso por medio de indicadores diseñados para explotar elementos del clima de los negocios. El informe recomienda mejoras institucionales y de comportamiento: diseñar una regulación y una tributación mejores, reducir las barreras a la competencia, mejorar los incentivos de los negocios, combatir la corrupción, fomentar la confianza y la legitimidad pública, y garantizar una ejecución apropiada de las regulaciones y las leyes.

Muchos de los nuevos temas pertinentes a la ayuda para el desarrollo incluyen a los asociados bilaterales y multilaterales para el desarrollo y la carga potencial de sus múltiples evaluaciones en los países en desarrollo. Su participación es la base del fundamento para las evaluaciones internacionales conjuntas, y tales evaluaciones pueden efectuarse a nivel de proyecto, país, sector o tema. Pueden producir eficiencias en costo y escala para las organizaciones para el desarrollo, y también armonización de métodos de evaluación que faciliten la comparación de los resultados.

Lo que sigue a continuación es una breve discusión de algunos de los principales impulsores de la agenda de desarrollo internacional y sus implicaciones para la evaluación. Estos impulsores incluyen los siguientes:

- Los ODM.
- El Consenso de Monterrey.
- La Declaración de París sobre la Eficacia de la Ayuda.
- La Iniciativa para los países pobres muy endeudados.
- El papel de las fundaciones.
- La prevención de conflictos y la reconstrucción posconflicto.
- La gobernanza.
- El combate al lavado de activos y la financiación del terrorismo.
- Las remesas de los trabajadores.
- Los problemas de género.
- El desarrollo del sector privado y el clima de inversiones.
- La sostenibilidad ambiental y social.
- Los bienes públicos globales.

Los autores anotan que la lista de impulsores está siempre en movimiento. Al salir este libro para su impresión, un impulsor clave que podría encabezar la lista es el de los efectos de la crisis financiera global en el mundo en desarrollo.

Los Objetivos de Desarrollo del Milenio (ODM)

El 20 de septiembre de 2000, 189 países integrantes de las Naciones Unidas y numerosas organizaciones internacionales adoptaron la Declaración del Milenio de las Naciones Unidas, a partir de la cual se derivaron, en parte, los ODM. Los ODM comprenden un conjunto de objetivos de desarrollo que debe cumplir la comunidad internacional para el año 2015, como resultado de la participación activa de los países, desarrollados y en desarrollo (recuadro 2.1). Estos ambiciosos objetivos están dirigidos a reducir la pobreza, al desarrollo humano y a crear asociaciones mundiales para lograr ambas cosas, y representan un alejamiento de la importancia que anteriormente había fijado la comunidad para el desarrollo en el crecimiento económico, que esperaba liberaría de la pobreza a los pueblos de la tierra. Los ODM se dirigen específicamente a una serie de medidas encaminadas a reducir la pobreza y mejorar las condiciones de vida de los pobres del mundo.

ODM: Conjunto de objetivos de desarrollo que debe cumplir la comunidad internacional para el año 2015.

Los ocho ODM incluyen un conjunto de 18 objetivos menos generales y 48 indicadores para medir el progreso. (Los países en desarrollo tienen distintas mezclas de estos 18 objetivos y distintas fechas para cumplirlos, dependiendo de su situación.) Los ODM son objetivos basados en resultados que deben medirse, controlarse y evaluarse, y plantean problemas importantes para los sistemas de evaluación en todos los países.

Muchos países en desarrollo carecen de capacidad para realizar SyE. Para llenar este vacío, las organizaciones de desarrollo han ofrecido formación de capacidad estadística y de SyE, asistencia técnica y apoyo.Los ODM están impulsando a los países en desarrollo a generar capacidad y sistemas de SyE, y se está acudiendo a las organizaciones de

Recuadro 2.1 *Los Objetivos de Desarrollo del Milenio*

- Erradicar la extrema pobreza y el hambre.
- Lograr la educación primaria universal.
- Promover la igualdad de los géneros y empoderar a la mujer.
- Reducir la mortalidad infantil.
- Mejorar la salud materna.
- Combatir el VIH/Sida, la malaria y otras enfermedades.
- Garantizar la sostenibilidad del medio ambiente.
- Desarrollar una asociación global para el desarrollo.

Fuente: http://www.unorg/milleniumgoals.

desarrollo para proveer asistencia técnica y financiación para estos esfuerzos. Muchos países en desarrollo se encuentran en las primeras etapas de formación de sistemas de SyE, y están abriéndose paso para construir sistemas basados en resultados que contribuirán a establecer el grado de cumplimiento de los ODM. La evaluación del éxito en el cumplimiento de los ODM requerirá el desarrollo y el uso eficaz de sistemas de evaluación. A su vez, el sistema de evaluación deberá integrarse a la arena política de los ODM, de modo que sea "claro para todos por qué es importante recopilar los datos, y cómo se utilizará la información para informar sobre los esfuerzos del gobierno y la sociedad civil para lograr los ODM" (Kusek, Rist y White, 2004, pp. 17-18).

Todos los años, el Banco Mundial y el Fondo Monetario Internacional (FMI) publican un *Informe sobre el Seguimiento Mundial* sobre los ODM. El informe proporciona un marco para la rendición de cuentas en la política para el desarrollo global.

El enfoque del *Informe sobre el Seguimiento Mundial 2004* fue sobre cómo le está yendo al mundo en la implementación de las políticas y acciones requeridas para cumplir los ODM y los efectos directos relacionados con el desarrollo. Destaca varias prioridades para fortalecer el ejercicio de control, las que incluyen:

- Fortalecer las estadísticas de desarrollo fundamentales, incluso mediante la implementación oportuna del plan de acción acordado por las agencias de estadísticas internacionales.
- Investigar los determinantes de los ODM, problemas críticos como la eficacia de la ayuda y el desarrollo de métricas más vigorosas para áreas clave de políticas, como la gobernanza y el impacto de las políticas de los países ricos en los países en desarrollo.
- Intensificar la colaboración con las agencias asociadas en este trabajo, aprovechando la ventaja comparativa, y garantizando que el enfoque del SyE sea coherente a través de las agencias (Banco Mundial, 2004b).

El *Informe sobre el Seguimiento Mundial 2005* señaló las oportunidades ofrecidas por un mejor desempeño económico en muchos países en desarrollo, y bosquejó una agenda de cinco puntos diseñada para acelerar el progreso:

- *Garantizar que los esfuerzos encaminados hacia el* desarrollo sean propios del país. Aumentar en forma gradual el impacto en el desarrollo por medio de estrategias de reducción de la pobreza propias del país y dirigidas por éste.
- *Mejorar el entorno para el crecimiento económico privado e impulsado por los sectores.* Fortalecer la gestión fiscal y la gobernanza, facilitar el entorno de los negocios y realizar inversiones en infraestructura.
- *Aumentar gradualmente la prestación de servicios sociales básicos.* Incrementar en forma rápida la oferta de trabajadores de salud y maestros, proveer mayor financiación y más flexible y predecible para estos servicios recurrentes e intensivos en costos, y fortalecer la capacidad institucional.
- *Desmantelar las barreras al comercio internacional.* Utilizar una Ronda Doha ambiciosa que incluya reformas principales de las políticas de comercio agrícola para desmantelar las barreras comerciales e incrementar la "ayuda para el comercio".
- Duplicar la ayuda para el desarrollo en los próximos cinco años. Además, mejorar la calidad de la ayuda, con un progreso más rápido en la coordinación y la armonización de la ayuda (Banco Mundial, 2005a).

El *Informe sobre el Seguimiento Mundial 2006* destacó el crecimiento económico, la mejor calidad de la ayuda, la reforma al comercio y la gobernanza como elementos esenciales para cumplir los ODM (Banco Mundial, 2006a). En el informe de 2007 se destacaron dos áreas temáticas clave: la igualdad de géneros y el enpoderamiento de la mujer (el tercer ODM) y los problemas especiales de estados frágiles, en donde se concentra en forma creciente la pobreza extrema (Banco Mundial, 2007f).

El Consenso de Monterrey

En marzo de 2002, representantes de los gobiernos de más de 170 países, incluyendo más de 50 jefes de Estado, se reunieron para discutir la propuesta del Consenso de Monterrey sobre Financiación para el Desarrollo. La propuesta reflejaba un intento de distribuir más dinero a los más pobres del mundo, que viven con menos de US$1 al día.

Para la evaluación para el desarrollo fue más significativo que el Consenso de Monterrey le diera mayor importancia a las responsabilidades en procura del cumplimiento de los ODM, e instó a los países en desarrollo a mejorar sus políticas y gobernanza, y a los países desarrollados a acelerar su apoyo, especialmente abriendo el acceso a sus mercados y suministrando más y mejor ayuda. En el documento se reconoció la necesidad de una mayor ayuda financiera para aumentar el nivel de vida de los países más pobres, pero no se fijaron objetivos para el incremento de la ayuda, el alivio de la mayor parte de las deudas, o la eliminación de barreras comerciales (Qureshi, 2004).

001El camino.indd 68 08/11/2010 03:27:42 p.m.

En un punto intermedio entre el año en que se adoptaron los ODM y su fecha límite de 2015, la Economic Commission for Africa publicó un informe de evaluación del progreso del continente en cuanto al cumplimiento de los compromisos con África para el Consenso de Monterrey. En el informe se concluyó que se había realizado un avance sustancial en el campo del alivio de la deuda externa, pero muy limitado en los demás campos principales del Consenso. Como asunto de interés para los evaluadores, en el informe se dice que:

> Existe el conocimiento de que el seguimiento de los compromisos contraídos por los países africanos y sus asociados en el desarrollo es esencial para el cumplimiento de los objetivos del Consenso de Monterrey. Los dirigentes africanos han reconocido esto y han implementado un mecanismo de seguimiento del progreso en la ejecución de sus compromisos y los de sus asociados en el desarrollo. La reciente institucionalización de una Conferencia Ministerial sobre Financiación para el Desarrollo constituye un paso audaz de los dirigentes africanos en este campo. La comunidad internacional ha implementado a su vez un mecanismo para seguir el desempeño de las entidades donantes. Por ejemplo, establecieron un Foro de Asociación Africana y un Panel sobre el Progreso Africano, mediante los cuales se seguirá el progreso en el cumplimiento de los compromisos clave sobre la financiación para el desarrollo. En último término, la eficacia de esos mecanismos de seguimiento se evaluará en cuanto a su capacidad de transformar las promesas de los asociados para el desarrollo en hechos concretos, pues solamente cumpliendo estos compromisos los países africanos y la comunidad internacional pueden reducir la pobreza en la región y establecer los fundamentos de un futuro más brillante para sus pueblos (Katjomulse y otros, 2007, p. vi).

La Declaración de París sobre la eficacia de la ayuda

La Declaración de París sobre la eficacia de la ayuda fue un acuerdo internacional para aumentar los esfuerzos en el manejo de la ayuda para los países en desarrollo. Más de 100 ministros, directores de agencias y otros funcionarios superiores firmaron este acuerdo el 2 de marzo de 2005.

Una característica importante para la evaluación de esta declaración fue el acuerdo de utilizar acciones controlables e indicadores como parte de la implementación del acuerdo. Se desarrollaron 12 indicadores para seguir y estimular el progreso en la obtención de ayuda más eficaz. Se fijaron objetivos para 11 de los 12 indicadores para el año 2010 (OCDE, 2005b).

Los indicadores y objetivos aprobados se organizan alrededor de cinco principios básicos:

- *Propiedad.* Los países asociados ejercen un liderazgo efectivo sobre sus políticas y estrategias de desarrollo y coordinan las acciones para el desarrollo.
- *Alineación.* Las organizaciones de desarrollo basan su apoyo general en las estrategias, instituciones y procedimientos de desarrollo, de los países asociados.

- *Armonización.* Las acciones de las organizaciones de desarrollo son más armonizadas, transparentes y eficaces colectivamente.
- *Administración por resultados.* Los gobiernos cambian su énfasis al manejo de los recursos y a una mejor toma de decisiones teniendo como base los resultados.
- *Responsabilidad mutua.* Las organizaciones de desarrollo y los asociados son responsables de los resultados sobre el desarrollo (Joint Progress toward Enhanced Aid Effectiveness, 2005).

En 2007, la OCDE publicó un informe decisivo con un resumen de los resultados de una encuesta base del estado de las cosas en 2005. En el informe (OCDE, 2007b) se presenta una evaluación de la ayuda tanto globalmente como para las organizaciones de desarrollo.

La OCDE realizó una encuesta para hacer seguimiento al progreso en el mejoramiento de la eficacia de la ayuda, según se había resaltado en el Consenso de Monterrey y más concretamente en la Declaración de París. Se sacaron las conclusiones siguientes:

- La Declaración de París aumentó la conciencia y promovió el diálogo a nivel de países sobre la necesidad de mejorar la provisión y el manejo de la ayuda.
- El ritmo del progreso en el cambio de actitud de las entidades donantes y sus prácticas sobre la gestión de la ayuda ha sido lento, y los costos transaccionales del suministro de la ayuda y su manejo continúan siendo muy altos.
- Es necesario fortalecer las estrategias para el desarrollo, mejorar el alineamiento del apoyo de los donantes con las prioridades internas, aumentar la credibilidad del presupuesto como herramienta para dirigir y asignar los recursos, e incrementar el grado de precisión en los estimados presupuestarios de los flujos de ayuda.
- Modificar la forma como se presta y maneja la ayuda implica nuevos costos que deben tener en cuenta las entidades donantes y los asociados.
- Los países y entidades donantes deberían utilizar marcos de evaluación del desempeño y presentación de información más efectivos en costos y orientados a resultados. Las entidades donantes deberían contribuir a generar capacidad y hacer un mayor uso de los sistemas de presentación de información de los países.
- Es necesario desarrollar sistemas de seguimiento más verosímiles para garantizar la responsabilidad mutua (Katjomulse y otros, 2007).

La iniciativa para los países pobres altamente endeudados (PPAE)

En 1996, el Banco Mundial y el FMI propusieron la Iniciativa PPAE, el primer enfoque integral para reducir la deuda externa de los países más pobres y más endeudados del mundo, y 180 países aprobaron la iniciativa.

La PPAE se diseñó para reducir las deudas a niveles sostenibles para los países pobres que buscan reformas de políticas económicas y sociales. Se utiliza cuando los mecanismos tradicionales de alivio de deuda no son suficientes para ayudar a los

países a salir del proceso de reprogramación de la deuda. La PPAE reduce el monto de la deuda, disminuye los pagos de servicio de la deuda y fomenta el gasto social.

La iniciativa incluye el alivio de deudas bilaterales y multilaterales. Se espera un recorte del servicio de la deuda externa para los países de la PPAE por unos US$50.000 millones. En enero de 2009 se habían aprobado paquetes de reducción de deuda para 34 países, 28 de ellos en África y 7 países adicionales fueron considerados elegibles para la ayuda (FMI, 2009).

La PPAE está vinculada a las estrategias integrales nacionales de reducción de la pobreza. En 1999, la comunidad para el desarrollo internacional acordó que los Documento(s) de Estrategias de Producción de la Pobreza (PRSP), deberían constituir la base para los préstamos de concesiones y el alivio de la deuda. Estas estrategias incluyen objetivos de desarrollo acordados para un período de 3 años, e incluyen también una matriz de políticas, un conjunto acompañante de indicadores mesurables y un sistema de SyE a través del cual se mida el progreso. Si un país cumple sus objetivos, se reduce su deuda, lo que proporciona incentivos para acelerar las reformas y aumentar la propiedad del país. Como condición para el alivio de la deuda, los gobiernos receptores deben tener capacidad de controlar, evaluar e informar sobre los esfuerzos de reforma y el progreso en la reducción de la pobreza. Esta condición creó una demanda de generación de capacidad y asistencia en SyE.

Algunos países en desarrollo, por ejemplo Uganda, han progresado en la evaluación y han calificado para más alivio proveniente de la PPAE. La falta de capacidad para evaluación ha constituido un problema para otros países participantes en la PPAE, por ejemplo, Albania, Madagascar y Tanzania. Estos países requieren ayuda adicional para desarrollar su capacidad de evaluación.

Al efectuar préstamos de concesión o garantías a países con niveles muy altos de deuda para mitigar el riesgo de futuras crisis de deuda, la PPAE sacó a relucir un nuevo tema de evaluación: ¿Cómo se evaluaría la eficacia de la subvención, frente a la de los préstamos, y con qué criterios? Esta pregunta crea nuevos problemas para los evaluadores para el desarrollo.

En septiembre de 2006, se cumplieron 10 años de la PPAE Initiative. Desde 1999, los gastos en reducción de la pobreza de la PPAE han aumentado, y los pagos de servicio de deuda han disminuido (Banco Mundial, 2007e), lo que sugiere que ha habido progreso en los resultados de la PPAE.

El papel de las fundaciones

En un estudio de la OCDE se estimó la cuantía que fundaciones filantrópicas han otorgado a países en desarrollo. Con el estudio:

> Se intentó efectuar un estimado serio de la cuantía de los fondos distribuidos por 15 de las más grandes fundaciones filantrópicas con donaciones internacionales, para 2002. El total fue de casi US$4.000 millones y el total de donaciones internacionales fue de unos US$2.000

> millones, lo que representa alrededor del 4% del total de la ayuda para el desarrollo y más o menos la mitad de las contribuciones que atribuye el Development Assistance Committee a… las ONG como un todo (grupo que incluye las fundaciones) (Oxford Analytica, 2004a).

El U.S. Council on Foundations comprende 56.000 fundaciones privadas y comunitarias en los Estados Unidos, y distribuye US$27.500 millones por año. El European Foundation Centre encontró unas 25.000 fundaciones en nueve países de la UE con gastos anuales de más de US$50.000 millones.

Varias fundaciones grandes dominan la escena global, y entre ellas se encuentran la Bill & Melinda Gates Foundation, la Ford Foundation, la Susan Thompson Buffet Foundation y la Soros Foundation/Open Society.

La red de Soros Foundation/Open Society Institute es un participante influyente en la escena para el desarrollo internacional, con programas en más de 50 países que ofrecen apoyo para educación, medios de comunicación, salud pública, la mujer, derechos humanos, arte y cultura y reformas sociales, económicas y legales (Soros Foundation Network, 2007).

Prevención de conflicto armado y reconstrucción posconflicto

En los años posteriores a la Guerra Fría, entre 1989 y 2001, hubo 56 conflictos armados importantes en 44 ubicaciones distintas. En 2003, se estimó que los conflictos afectaban a más de 1.000 millones de personas. Durante este período la mayoría de los conflictos duraron siete años o más. Los costos totales de las guerras civiles en particular son enormes. "Creando territorio fuera del control de cualquier gobierno reconocido, los conflictos armados fomentan el tráfico de drogas, el terrorismo y la propagación de las enfermedades" (Collier y otros, 2003).

La pobreza es tanto una causa como una consecuencia del conflicto. En 16 de los 20 países más pobres se ha experimentado una guerra civil importante. En promedio, los países que emergen de la guerra tienen un 44% de probabilidad de reincidir en los primeros cinco años de paz.

El tratamiento de la reconstrucción posconflicto implica coordinar grandes números de organizaciones de desarrollo bilateral y multilateral. Por ejemplo, en Bosnia-Herzegovina estuvieron activas 60 organizaciones de desarrollo, en Cisjordania y la Franja de Gaza, 50, y en Afganistán, 82. La reconstrucción posconflicto ocasiona tensiones en los mecanismos de coordinación de ayuda para garantizar que se satisfagan las necesidades y evitar vacíos en la ayuda.

La reconstrucción posconflicto implica algo más que simplemente reconstruir infraestructura, y a menudo es necesario proveer apoyo para reconstruir las instituciones, la democracia, las elecciones, ONG, la sociedad civil, las fuerzas de policía, y los costos presupuestarios de iniciación y gastos, el alivio de deuda, el apoyo para la balanza de pagos, la igualdad de géneros, el desmonte de minas, el restablecimiento de refugiados y desplazados internos y la desmovilización y la reintegración de ex combatientes.

Debido a las preocupaciones sobre la corrupción y la necesidad de apalancar la ayuda oficial al desarrollo, la reconstrucción posconflicto a menudo ha implicado crear nuevos instrumentos y mecanismos de endeudamiento. En Cisjordania y la Franja de Gaza, por ejemplo, se creó un fondo de fideicomiso de organizaciones de desarrollo multilateral para dar soporte a los gastos presupuestarios de puesta en marcha y gastos para la nueva administración palestina. Dichos instrumentos y mecanismos ya son comunes en regiones de posconflicto en otras partes del mundo.

Los programas de reconstrucción posconflicto (programas multisectoriales que cuestan miles de millones de dólares) incrementan los niveles de dificultad y de escala de la evaluación (Kreimer y otros, 1998). Los evaluadores deben examinar el impacto que causan tales enfoques de desarrollo con gastos muy concentrados al inicio en la reconstrucción y la reconciliación posconflicto. Evaluar un proceso de coordinación que reúne un grupo grande y diverso de entidades de apoyo bilaterales, multilaterales y otros plantea nuevos problemas.

Los evaluadores deben examinar nuevos proyectos y programas en campos no tradicionales de ayuda para el desarrollo, como la desmovilización y la reintegración de ex combatientes, y el desmonte de minas de la tierra. Deben evaluar también nuevos tipos de mecanismos de organización de desarrollo e instrumentos de endeudamiento, como fondos fiduciarios de organizaciones de desarrollo multilateral.

Estas organizaciones analizan cada vez más las causas económicas y las consecuencias de los conflictos y buscan formas de impedirlos. Hay un mayor interés en las comunidades y relaciones sociales, en los grupos étnicos y religiosos; la gobernanza y las instituciones políticas; los derechos humanos; la seguridad; las estructuras y el desempeño económico; el medio ambiente y los recursos naturales, y los factores externos. Esto significa que los evaluadores deben analizar lo que se está haciendo, y lo que podría hacerse en el contexto del desarrollo para impedir que surjan los conflictos.

Gobernanza

Pese a que con mucha frecuencia se reconocía tras puertas cerradas, el problema de la gobernanza y la corrupción pasó al primer plano del desarrollo apenas a mediados de los años noventa, y desde entonces se han firmado convenciones internacionales para atender el problema de la corrupción alrededor del mundo. Las Naciones Unidas y la OCDE han adoptado convenciones sobre la corrupción que incluyen previsiones sobre prevención y criminalización de la corrupción, cooperación internacional para la recuperación de activos y medidas contra el soborno.

Los bancos de desarrollo multilateral han instituido también programas anticorrupción, y se efectúan préstamos para ayudar a los países a formar instituciones eficientes y responsables en el sector público. La medidas de gobernanza y anticorrupción se incluyen en las estrategias de ayuda a los países. Los programas de gobernanza buscan promover:

- La anticorrupción.
- El buen manejo del gasto público.
- La reforma del servicio civil.
- La reforma judicial.
- La administración, descentralización, el *e-government* (el uso de las TIC por parte del gobierno) y la prestación de servicios públicos.

Transparencia Internacional (TI), una ONG cuyo propósito es colocar la "corrupción en la agenda global", fue creada y puesta en marcha a principios de los años noventa y actualmente tiene capítulos en 88 países. Trabaja con asociados locales, nacionales, regionales e internacionales (gubernamentales y no gubernamentales) para combatir la corrupción (http://www.transparency.org/).

El Índice de Percepción de Corrupción anual de TI clasifica unos 140 países con base en la percepción de los funcionarios públicos de la corrupción. Su Índice de Pagadores de Soborno anual clasifica los países exportadores sobre la base de la incidencia de sobornos.

Algunos estimados informan que más de US$1.000 millones se pierden cada año por la corrupción. Medir la corrupción y los costos de ésta ha constituido un verdadero problema para la comunidad internacional, pero la

> creciente disponibilidad de encuestas y sondeos de muchas instituciones, con datos sobre distintas dimensiones de gobernanza, ha permitido construir un banco mundial de datos sobre ésta. Utilizando varias decenas de fuentes y variables diferentes, y también como técnica novedosa de agregación, el banco de datos cubre ya 200 países en el mundo, y contiene indicadores agregados en campos como el cumplimiento de la ley, la corrupción, la calidad reguladora, la eficacia del gobierno, la voz y la rendición de cuentas, y la inestabilidad política (Banco Mundial, 2007c, p. 1).

Las organizaciones de desarrollo y los evaluadores pueden utilizar estos datos como medida de la eficacia de la ayuda, y los hallazgos sugieren que en donde es más alta la corrupción, es proporcionalmente más alta la posibilidad de que la ayuda se malgaste. La gestión basada en resultados se utiliza para identificar y controlar los determinantes y las instituciones más vulnerables en la estructura de gobernanza de un país. Los datos contribuyen a desmitificar y tratar en forma más objetiva los problemas de gobernanza que anteriormente se encontraban oscurecidos. Los datos generados ayudarán también a los evaluadores a compilar más hallazgos de evaluación cuantitativa relativos a la adquisición de enseñanzas. Al mismo tiempo, la evaluación de climas de inversión y entornos de negocios implicará conceptos difíciles y espinosos (*véase* más adelante la sección sobre desarrollo del sector privado y el clima de inversión).

Este nuevo campo está evolucionando con rapidez y requerirá que los evaluadores atiendan los nuevos desarrollos y datos en forma oportuna.

Financiación contra el lavado de activos y el terrorismo

La financiación contra el lavado de activos y el terrorismo forman parte del paisaje general de la anticorrupción.

El *lavado de activos* es la práctica de efectuar transacciones financieras para ocultar la identidad, la fuente y el destino del dinero en cuestión. En el pasado, el término "lavado de activos" sólo se aplicaba a las transacciones financieras relacionadas con actividad de otras formas criminales. Hoy su definición se extiende a menudo a los reguladores del gobierno, como la Comisión de Valores y Cambios (SEC), para incluir cualquier transacción financiera no transparente con base en la ley. Como resultado, la actividad ilegal de lavado de dinero es practicada comúnmente por individuos comunes, pequeñas y grandes empresas, funcionarios corruptos y miembros del crimen organizado, como traficantes de drogas o miembros de la mafia (InvestorDictionary.com, 2006).

Lavado de activos: La práctica de efectuar transacciones financieras para ocultar la identidad, la fuente y el destino del dinero en cuestión.

Con un estimado de US$1.000 millones (2 a 5% del PIB mundial) lavados cada año, según el FMI, el lavado de dinero constituye un grave problema internacional en crecimiento, que afecta a los países en desarrollo y desarrollados por igual (Camdessus, 1998, p. 1). La globalización y la apertura o la distensión de las fronteras han facilitado las actividades criminales transnacionales y los consecuentes flujos financieros ilegales. Las iniciativas globales contra el lavado de dinero han adquirido nueva importancia con la propagación del terrorismo.

El lavado de dinero puede costarles muy caro a los países en desarrollo.

> Los mercados financieros emergentes y los países en desarrollo constituyen… objetivos importantes y víctimas fáciles para los lavadores de dinero, que continuamente buscan nuevos lugares y formas de evitar el ojo vigilante de la ley. Las consecuencias de operaciones de lavado de dinero causan estragos particularmente a las economías en desarrollo. Sin vigilancia alguna, los lavadores de dinero pueden manipular los sistemas financieros del país para operar y ampliar sus actividades ilícitas… y rápidamente socavar la estabilidad y el desarrollo de las instituciones establecidas (IFAC, 2004, p. 5).

En 1989 el Grupo de los Siete (G-7) creó el Grupo de Acción Financiera contra el Blanqueo de Capitales (GAFI) de la OCDE, la cual incluye ya 31 países y territorios miembros y 2 organizaciones regionales. Este organismo intergubernamental de formulación de políticas aspira a desarrollar y promover políticas nacionales e internacionales para combatir el lavado de dinero y la financiación del terrorismo.

El control y la evaluación de la implementación forman parte del mandato de la GAFI y se realiza multilateralmente, por revisión de pares, y por evaluación mutua. El proceso de SyE implica que:

> Cada país miembro es examinado a su turno por la GAFI con una visita en el sitio que realiza un equipo de tres o cuatro expertos seleccionados, en el campo legal, financiero y del cumplimiento de la ley, de otros países miembros. El propósito de la visita es elaborar un

> informe en el que se presenta una evaluación del grado al que el país evaluado ha avanzado en la implementación de un sistema eficaz contra el lavado de dinero y destacar las áreas en las que puede requerirse todavía progreso adicional (GAFI, 2007, p. 1).

La GAFI ha establecido una serie de medidas que deben tomarse en caso de incumplimiento.

Remesas de trabajadores

Las remesas mundiales anuales enviadas por trabajadores emigrantes a su país de origen sobrepasaron el ritmo de la ayuda oficial para el desarrollo. Las remesas anuales aumentaron de US$60.000 millones en 1998 a US$80.000 millones en 2002, y se estimaron en US$100.000 millones para 2003. Estas cifras superan los US$50.000 a US$60.000 millones anuales de la ayuda oficial para el desarrollo, y son comparables a los US$143.000 millones en flujos de capital privado en 2002. Las remesas suelen ser más estables que los flujos de capital privado (Banco Mundial, 2003; Oxford Analytica, 2004b).

Las remesas mundiales causan un fuerte impacto en la reducción de la pobreza. "En promedio, un aumento del 10% en la proporción de emigrantes internacionales en la población de un país lleva a una disminución del 1,9% en la proporción de personas que viven en la pobreza (US$1/persona/día)" (Adams y Page, 2003, p. 1). Las remesas mundiales contribuyen a financiar el consumo local en vivienda, agricultura, industria y la creación de nuevas pequeñas y medianas empresas en el país receptor.

Los países desarrollados y en desarrollo y las organizaciones conocen estas tendencias, y buscan formas de capitalizar estos flujos hacia inversiones. Un Summit Plan del Grupo de los Ocho (G-8) exhortó a sus miembros y a los gobiernos de los países en desarrollo a

> Facilitar los flujos de remesas de las comunidades en el exterior para ayudar a las familias y a los pequeños empresarios [negocios], incluso estimulando la reducción de los costos de las transferencias de remesas y la creación de fondos locales de desarrollo para inversiones productivas, mejorando el acceso de los receptores de las remesas a los servicios financieros y mejorando la coordinación (G-8, 2004a, p. 7).

Es probable que las remesas efectuadas a través del sistema bancario aumenten, según aumenten las restricciones a las transferencias informales, a causa de regulaciones de control más cuidadosas para represar la financiación a organizaciones terroristas mediante mecanismos informales (ver adelante) y una reducción en las tarifas bancarias como resultado de la mayor competencia en el sector para capturar el mercado mundial de remesas.

El impacto de las remesas en los países en desarrollo no se ha articulado ni probado totalmente. Seguir la pista de las remesas mundiales y verterlas a nuevos tipos de inversiones y fondos plantea nuevos problemas a los evaluadores. Como los profesionales para el desarrollo no han diseñado todavía formas de capturar las remesas y apalancarlas para reducir la pobreza, los evaluadores están atentos a esta área con gran interés.

Género: desde la mujer en el desarrollo hasta el género y el desarrollo, y la incorporación del género a las actividades habituales

El género se refiere a los papeles socialmente atribuidos a las mujeres y a los hombres. El *análisis de género* examina el acceso y el control sobre los recursos por parte de hombres y mujeres. Se refiere también a una manera sistemática de determinar las necesidades y preferencias con mucha frecuencia distintas de hombres y mujeres, y a los diferentes efectos del desarrollo en mujeres y hombres. El análisis de género tiene en cuenta cómo interactúan la clase, raza, grupo étnico y otros factores con el género, para producir resultados discriminatorios. El análisis de género se ha dirigido tradicionalmente hacia las mujeres por el vacío existente entre hombres y mujeres en términos de cómo se benefician de la educación, el empleo, los servicios, etc.

Análisis de género: Examen del acceso y el control sobre los recursos por parte de hombres y mujeres.

Las mujeres representan un poco más de la mitad de la población mundial y su papel es clave en el desarrollo económico. Sin embargo, no se ha realizado aún su potencial completo de participación en el desarrollo socioeconómico. Es un hecho que las mujeres y los niños representan la mayoría de los pobres del mundo.

> Las mujeres producen la mitad de los alimentos en algunas partes del mundo en desarrollo, tienen la mayor parte de la responsabilidad por la seguridad alimentaria del hogar y representan una cuarta parte de la fuerza laboral en la industria y una tercera parte en los servicios… Sin embargo, debido a la mayor limitación en el acceso a la educación y otras oportunidades, la productividad de la mujer en relación con la del hombre es todavía baja. El mejoramiento de la productividad de la mujer contribuye al crecimiento, la eficiencia y la reducción de la pobreza, que son objetivos clave para el desarrollo en todas partes (Banco Mundial, 1994, p. 9).

Las tendencias recientes sobre el papel de la mujer en el desarrollo han evolucionado, alejándose del tradicional enfoque de "mujeres en el desarrollo" (WID, por sus iniciales en inglés) y del "género y desarrollo" (GAD, por sus iniciales en inglés) y acercándose a una enfoque más integral de "incorporación del género a las actividades habituales". La estrategia WID se centraba en la mujer como objetivo o grupo de interés especial de beneficiarios en proyectos, programas y políticas. "La WID reconoce que la mujer es participante activa, aunque no se le ha reconocido como tal, en el proceso de desarrollo, aportando una contribución crítica al crecimiento económico… como recurso no explotado; la mujer debe introducirse en el proceso para el desarrollo" (Moser, 1995, p. 107).

El enfoque GAD se centra en las fuerzas sociales, económicas, políticas y culturales que determinan cómo participan, se benefician y se controlan los hombres y las mujeres en los recursos y las actividades de los proyectos. Destaca a menudo las diferentes necesidades y preferencias de mujeres y hombres. Este enfoque cambia la posición de las mujeres como un grupo a las relaciones entre hombres y mujeres determinadas socialmente.

El progreso en la igualdad de géneros y el apoderamiento de la mujer están incorporados en los ODM, que incluyen objetivos e indicadores para medir y evaluar el progreso. El Comité de Asistencia para el Desarrollo (CAD) de la OCDE produjo también preguntas de orientación para asistir a los administradores en la evaluación de las actividades para el desarrollo. Las preguntas incluyen las siguientes:

- ¿Ha tenido éxito el proyecto en promover igualdad de oportunidades y beneficios para hombres y mujeres?
- ¿Han tenido ventaja o desventaja los hombres o las mujeres por el proyecto?
- ¿Ha sido eficaz el proyecto en integrar los géneros a la actividad de desarrollo? (Woroniuk y Schalkwyk, 1998).

El presupuesto por géneros es una forma de implementar y evaluar cuánto del presupuesto nacional beneficia a hombres y mujeres. Otra forma de medir y evaluar la ayuda es la de examinar el grado al cual la ayuda para el desarrollo beneficia sectores que "incluyen a la mujer, ayudan a la mujer, empoderan a la mujer y generan resultados para la mujer" (Jalan, 2000, p. 75). Dada la importancia actual que se concede a los enfoques integrales y las asociaciones, debe efectuarse, integrarse y coordinarse también la evaluación de políticas de incorporación de género dentro de, y entre países, organizaciones y agencias asociadas para el desarrollo. En toda evaluación es importante tener en cuenta cómo afecta el proyecto, programa o política, en forma diferencial a hombres y mujeres.

Desarrollo del sector privado y clima de inversiones

Bajo el encabezado de desarrollo del sector privado y clima de inversiones entran muchos asuntos, algunos de ellos son el papel del sector privado y la inversión extranjera directa (IED) en la reducción de la pobreza, la privatización, la participación privada en los servicios de infraestructura y las asociaciones público-privadas, y la creación y apoyo a micro y pequeñas y medianas empresas, a través de intermediarios fiscales.

Inversión del sector privado

Se ha reconocido de manera creciente que la inversión del sector privado es crítica para reducir la pobreza en el mundo en desarrollo. En 1990, la inversión del sector privado en los países en desarrollo fue de unos US$30.000 millones al año, mientras la ayuda para el desarrollo llegó a unos US$60.000 millones. Para 1997, la inversión del sector privado en los países en desarrollo había llegado a US$300.000 millones, y la ayuda para el desarrollo había disminuido a US$50.000 millones. El desarrollo del sector privado creció así de la mitad del tamaño de la ayuda para el desarrollo hasta seis veces su tamaño en menos de 10 años.

Ayuda oficial para el desarrollo

Una medida del clima de inversiones es la *ayuda oficial para el desarrollo* (AOD), que el *Glossary of Statistical Terms* de la OCDE (2002a) define como:

> Flujos de financiación oficial administrados con el principal objetivo de la promoción para el desarrollo económico y el bienestar de los países en desarrollo, con carácter de concesión y con un elemento de subvención de al menos el 25% (utilizando una tasa fija del 10% de descuento). Por convención, los flujos de AOD comprenden contribuciones de agencias del gobierno donantes, en todos los niveles, a países en desarrollo (AOD bilateral) y a instituciones multilaterales. Los recibos de AOD comprenden desembolsos de donantes bilaterales e instituciones multilaterales. (De una página web identificada como Ayuda Oficial para el Desarrollo).

Ayuda oficial para el desarrollo: Flujos de financiación oficial administrados a tasas de concesión para promover el desarrollo económico y el bienestar de los países en desarrollo.

En 1997, los niveles de ayuda subieron antes de alcanzar un límite que continuó hasta 2001. La AOD total de los miembros del CAD subió un 7% en 2001 y un 5% en 2003. En 2005, la ayuda oficial para el desarrollo de los miembros del CAD subió un 32%, en gran parte como resultado del incremento en la ayuda por el tsunami de 2004 y el alivio de la deuda para Irak y Nigeria (OCDE, 2005a). En 2006, la AOD bajó un 4,6% como resultado de la deuda excepcionalmente alta y la ayuda humanitaria en 2004 (OCDE, 2006).

La AOD creció de manera continua en la última década, y se espera que continúe creciendo, pues las entidades donantes se han comprometido a incrementarla gradualmente de manera significativa para cumplir los ODM. Para utilizar en forma efectiva dicho incremento a nivel de países, los donantes y receptores deben resolver varios problemas de implementación, en particular:

- Lograr la complementariedad en las prioridades y programas de desarrollo nacionales, regionales y globales.
- Fortalecer la capacidad de los países receptores para hacer un uso efectivo de la AOD potencialmente incrementada de rápido desembolso, por ejemplo, con apoyo presupuestario (Banco Mundial, 2007a).

Inversión extranjera directa (IED)

Otra medida del clima de inversiones es la IED, que tiene un papel extraordinario y creciente en los negocios mundiales. La IED es una inversión en otros países por parte de un inversionista de por lo menos un 10% del capital en acciones de una empresa, para establecer una participación financiera duradera en la empresa y ejercer algún grado

Inversión extranjera directa: Inversión a través de las fronteras de por lo menos un 10% del capital en acciones de un país, para establecer una participación financiera duradera en la empresa y ejercer algún grado de influencia en sus operaciones.

de influencia en sus operaciones. Es citada a menudo como un impulsor importante del crecimiento económico, y se considera que aporta determinados beneficios a las economías nacionales (InvestorDictionary.com, 2006).

El mayor incremento en la IED entre los años setenta y 1999 ocurrió en los países en desarrollo, en los que los flujos anuales aumentaron de un promedio de menos de US$10.000 millones en la década de 1970 y menos de US$20.000 millones en la de 1980, a US$179.000 en 1998 y US$208.000 millones en 1999. Estos flujos representaron una gran parte de la IED mundial (Graham y Spaulding, 2005).

La Conferencia de las Naciones Unidas para el Comercio y el Desarrollo (UNCTAD) (2008) informó un crecimiento mundial de la IED en 2007. El informe documenta un aumento en la IED en los tres grupos de economías: países desarrollados, economías en desarrollo y Europa sudoriental y la Comunidad de Estados Independientes (CEI). Los resultados reflejan la propensión a un alto crecimiento de las corporaciones transnacionales y el fuerte desempeño económico en muchas partes del mundo, al igual que mayores beneficios corporativos y abundancia de dinero, que elevó el valor de las fusiones y adquisiciones internacionales. Tales transacciones constituyeron una gran porción de los flujos de IED en 2007, si bien el valor de las fusiones y adquisiciones disminuyó en el segundo semestre del año.

Los flujos de IED hacia los países desarrollados subieron por cuarto año consecutivo en 2007, llegando a un quintillón de dólares. Los flujos hacia los países en desarrollo y economías de transición en Europa suroriental y la CEI subieron un 16% y un 41% respectivamente, superando niveles históricos.

Privatización

La privatización de empresas de propiedad del Estado constituyó una tendencia particularmente fuerte en los años noventa, cuando muchos países buscaron el paso de economías socialistas a economías orientadas al mercado. Aún es una fuerza importante en muchos países donde el Estado todavía posee y opera muchos activos económicos.

"Más de 100 países, en todos los continentes, han privatizado algunas o la mayor parte de las empresas estatales, en todos los sectores concebibles de infraestructura, manufactura y servicios… Un estimado de 75.000 empresas medianas y grandes han sido privatizadas alrededor del mundo, junto a cientos de miles de pequeñas unidades de negocios…" El producto total generado se estima en más de US$735.000 millones (Nellis, 1999). La privatización es controvertida y el debate sobre si, cuándo y cómo privatizar continúa. No constituye una panacea para las enfermedades económicas, pero ha demostrado ser una herramienta útil para promover ganancias netas de bienestar y mejores servicios para la economía y la sociedad.

001El camino.indd 80 08/11/2010 03:27:43 p.m.

Implicaciones para la evaluación

¿Cómo ha respondido la comunidad de la evaluación para el desarrollo a estas iniciativas? La Corporación Financiera Internacional (CFI) evalúa los efectos en las intervenciones a nivel de proyectos, y utiliza Instantáneas de Entorno de los Negocios (EN) para "presentar indicadores mesurables en una variedad de temas del entorno de los negocios y a través del tiempo" (CFI, 2007). Con esta nueva herramienta se compilan datos dispares, indicadores e información de proyectos sobre el entorno de los negocios para un país y son accesibles en un formato consistente. Los profesionales para el desarrollo y los autores de políticas pueden utilizar las Instantáneas EN para obtener un cuadro integral del entorno de los negocios en países particulares. Las Instantáneas EN pueden utilizarse también como herramienta de seguimiento o de planificación.

¿Cómo evaluar esta clase de actividades? A nivel general, pueden examinarse cuatro indicadores posibles, a saber:

- Desempeño de los negocios.
- Sostenibilidad económica.
- Efectos ambientales.
- Desarrollo del sector privado.

El *Informe sobre el Desarrollo Mundial 2005* resalta encuestas sobre el clima de inversiones, el entorno de los negocios y el desempeño de las empresas, que puede ayudar a identificar cómo pueden mejorar los gobiernos el clima de las inversiones para empresas de todo tipo. Las encuestas cubrieron 26.000 empresas en 53 países en desarrollo y 3.000 microempresas y empresas informales en 11 países.

Estas encuestas permiten la comparación de condiciones existentes y las pruebas de comparación de condiciones para hacer un seguimiento de los cambios en el tiempo. El instrumento de la encuesta es un conjunto central de preguntas, y varios módulos que pueden utilizarse para explorar con mayor profundidad aspectos específicos del clima de inversiones y su relación con la productividad a nivel de empresas.

Las preguntas pueden clasificarse en tres grupos:

- Las que generan información para perfil de negocios.
- Las utilizadas para el perfil del clima de inversiones en el que operan las empresas.
- Las que generan indicadores de desempeño firme.

Los indicadores utilizados fueron los siguientes:

- Incertidumbre política (principal restricción: interpretación impredecible de las regulaciones).
- Corrupción (principal restricción: los sobornos).
- Cortes (principal restricción: falta de confianza de que las cortes mantengan el derecho de propiedad).

- Crimen (principal restricción: pérdidas por el crimen). Otras fuentes de indicadores del clima de inversiones incluyeron un servicio de riesgo de los negocios, calificación de crédito de los países (Euromoney Institutional Investor), indicadores de riesgo de los países (World Markets Research Center), un servicio sobre riesgo de los países (Economist Intelligence Unit) y el *Global Competitiveness Report* (Foro Económico Mundial).

Los bancos de desarrollo multilateral, las instituciones financieras internacionales, las organizaciones para el desarrollo y el sector privado, participaron todas en dichas encuestas, suministrando información y asesoría valiosa. También se realizan evaluaciones continuas y periódicas y evaluaciones sobre el clima de inversiones en países de todo el mundo. Un ejemplo notorio es el de la base de datos de *Doing Business,* que presenta medidas objetivas de las regulaciones de negocios y su cumplimiento. Los indicadores comparables en 145 economías muestran los costos regulatorios de los negocios, y estos indicadores pueden utilizarse para analizar regulaciones específicas que estimulan o restringen la inversión, productividad y crecimiento (Banco Mundial, 2007d). Sin embargo, un informe del Independent Evaluation Group del Banco Mundial señaló que se necesitaban mejoras a los indicadores de Doing Business (IEG, 2008).

Sostenibilidad ambiental y social

La *responsabilidad social de las empresas* (RSE) implica tener en cuenta activamente el impacto económico, ambiental y social y las consecuencias de las actividades de las empresas. Las empresas del sector privado, las organizaciones y los gobiernos están a la búsqueda de nuevas formas de garantizar que las actividades y los servicios de las empresas no perjudiquen la economía, la sociedad o el medio ambiente en los países y sectores en donde operan. Por ejemplo, el gobierno británico adoptó varias políticas y legislación para estimular la RSE en general y la sostenibilidad ambiental y social en particular.

Responsabilidad social de las empresas: Aceptación de las empresas privadas de la necesidad de tener en cuenta el impacto económico, ambiental y social y las consecuencias de sus actividades.

> El gobierno considera la RSE como la contribución de las empresas a nuestros objetivos de desarrollo sostenible. En esencia se trata de cómo las empresas tienen en cuenta el impacto económico, social y ambiental en la forma como operan, maximizando los beneficios y minimizando los aspectos negativos… El enfoque del gobierno es estimular e incentivar la adopción de la RSE y la presentación de información sobre ella mediante orientaciones de buenas prácticas y, en lo que sea apropiado, la regulación inteligente y los incentivos fiscales. (BEER, 2004, par. 1).

Un ejemplo de esfuerzo internacional para la sostenibilidad ambiental y social es el de los *Equator Principles,* suscritos por 10 instituciones financieras de Occidente en 2003. Los principios fueron desarrollados por bancos del sector privado, y representan un enfoque de la industria para determinar, evaluar y manejar los riesgos ambientales

y sociales en la financiación de proyectos del sector privado. En 2006, se adoptó una versión corregida de los Equator Principles, la que refleja las revisiones a los *Performance Standards on Social and Environmental Sustainability* de la CFI. La versión de 2006 de los Equator Principles se aplica a todos los países y sectores y a la financiación de todos los proyectos con costos de capital que superen los US$10 millones.

La CFI y 61 bancos comerciales importantes (en Norteamérica, Europa, Japón y Australia) adoptaron en forma voluntaria los Equator Principles para financiar proyectos en todo el mundo. Las entidades buscan asegurar que los proyectos que financian se desarrollen de una forma socialmente responsable y reflejen prácticas de administración ambiental responsables. El propósito de los Equator Principles es servir como base y marco común para la implementación de procedimientos y estándares ambientales y sociales individuales e internos para las actividades de financiación de proyectos en todos los sectores industriales globalmente. Al adoptar estos principios, las entidades se comprometen a revisar cuidadosamente todas las propuestas para las que sus clientes solicitan financiación de proyectos, y se comprometen a no efectuar préstamos directamente a proyectos si el prestatario no cumple o no puede cumplir sus políticas y procesos ambientales y sociales. Los estándares cubren aspectos ambientales, de salud y seguridad, pueblos indígenas, hábitat natural y reasentamiento (Equator Principles, 2007).

Pero hacer un compromiso público con los principios y aplicarlos de buena fe son cosas bien distintas. La BankTrack, una red de 18 ONG internacionales especializadas en el sector financiero, ha tenido un papel importante en su contribución para el control de la manera en que se practican los Equator Principles. Es crítica de los requerimientos de información de los principios y la forma como hace seguimiento a las entidades financieras (BankTrack, 2008).

Bienes públicos globales

Los economistas definen los *bienes privados* como aquellos bienes cuyo consumo por parte de una persona reduce la cantidad disponible para otras, al menos hasta que se produzca mayor cantidad (Linux Information Project, 2006). Los bienes privados suelen ser artículos tangibles.

Los economistas definen los *bienes públicos* como los productos que pueden consumir los individuos tanto como deseen, sin reducir la cantidad disponible para otras (Linux Information Project, 2006). Por ejemplo, el aire limpio es un bien público porque respirarlo no reduce la cantidad de aire limpio disponible para los demás. Los bienes públicos suelen ser artículos intangibles y muchos pertenecen a la categoría de información o conocimiento.

Los *bienes públicos globales* son los bienes públicos que afectan a todo el mundo. Ejemplos de bienes públicos globales son los derechos de propiedad, la seguridad, la

Bienes privados: Bienes cuyo consumo por parte de una persona reduce la cantidad disponible para otras.

Bienes públicos: Bienes cuyo consumo por parte de una persona no reduce la cantidad disponible para otras.

Bienes públicos globales: Bienes públicos no rivales y no excluibles en todo el mundo, en lugar de solo dentro de las fronteras de un país.

estabilidad financiera y un medio ambiente limpio. En realidad, la evaluación para el desarrollo puede considerarse como una clase de bien público porque se extiende más allá de las fronteras de cualquier organización individual. Un buen estudio de evaluación puede tener efectos positivos en toda la comunidad de desarrollo (Picciotto y Rist, 1995, p. 23).

Los bienes públicos globales son importantes porque, con la mayor apertura de las fronteras nacionales, el dominio público de los países se ha entrelazado. Un bien público en un país depende con mucha frecuencia de la política interna y de hechos y elecciones de políticas de otros países o internacionales (gpgNet, 2008). Todas las personas dependen de bienes públicos y ni los mercados ni el individuo más rico pueden prescindir de ellos.

La evaluación está en gran parte ausente a nivel global:

> Los programas de colaboración diseñados para proporcionar bienes públicos globales no están sujetos a la evaluación *ex ante* independiente y, como resultado, a menudo carecen de objetivos claros y de indicadores de desempeño verificables. Además, el impacto de las políticas de los países desarrollados en los países pobres no se evalúa sistemáticamente aun cuando la ayuda, la deuda, la inversión extranjera, la polución, los patrones de migración y los sistemas de propiedad intelectual, se configuran por las decisiones de los gobiernos de los países desarrollados (Picciotto, 2002b, p. 520).

El control de la propagación del VIH/Sida y su eliminación constituye otro ejemplo de bien público global que se encuentra a la cabeza de muchas agendas internacionales. El impacto de la globalización en los pobres no se ha evaluado aún. La evaluación para el desarrollo debe ser más nativa, más global y más transnacional" (Chelimsky y Shadish, 1997).

En 2004, el Independent Evaluation Group del Banco Mundial publicó una evaluación de la participación del banco en los programas globales. Para el informe, titulado *Evaluating the World Bank's Approach to Global Programs: Addressing the Challenges of Globalization,* se investigaron 26 programas globales apoyados por el banco, y se adquirieron enseñanzas sobre el diseño, la ejecución y la evaluación de los programas globales (Banco Mundial, 2004a). En el informe se destacan 5 de sus 18 hallazgos:

- La estrategia del banco para programas globales está definida en forma deficiente.
- Los programas globales han incrementado muy poco la ayuda general.
- La voz de los países en desarrollo no está representada adecuadamente.
- Los programas globales revelan vacíos en inversión y en políticas públicas globales.
- Se requiere supervisión de los programas globales independientes.

En el informe se hacen las recomendaciones siguientes:

- Establecer un marco estratégico para la participación del banco en los programas globales.

- Vincular la financiación con las prioridades.
- Mejorar la selectividad y la supervisión del portafolio de programas globales.
- Mejorar la gobernanza y la gestión de los programas individuales.
- Efectuar evaluación adicional.

Resumen

Los países han adoptado varios enfoques para establecer sistemas de evaluación. El enfoque que abarca la totalidad del gobierno incluye el establecimiento de base amplia, e integral, del sistema en todo el gobierno. El enfoque circunscrito se concentra en una parte o sector del gobierno. En el enfoque mixto, algunas partes o sectores del gobierno se evalúan en forma integral y otras reciben un tratamiento más esporádico.

En los países en desarrollo es más difícil crear un sistema de evaluación, porque con mucha frecuencia carecen de sistemas políticos democráticos, tradiciones empíricas fuertes, funcionarios públicos capacitados en ciencias sociales y sistemas e instituciones administrativas eficientes. Las organizaciones de desarrollo se centran en ayudar a los países en desarrollo a adquirir la capacidad de crear y mantener sistemas de evaluación.

La evaluación recibe la influencia de muchos temas complejos en el desarrollo. Algunos de los principales impulsores de la agenda para el desarrollo son:

- Los ODM.
- El Consenso de Monterrey.
- La Declaración de París.
- La Declaración de París sobre la Eficacia de la Ayuda.
- La Iniciativa PPAE.
- El papel de las fundaciones
- La prevención de conflictos y la reconstrucción posconflicto.
- La gobernanza.
- El combate contra el lavado de activos y la financiación del terrorismo.
- Las remesas de los trabajadores.
- Los problemas de género.
- El desarrollo del sector privado y el clima de inversiones.
- La sostenibilidad ambiental y social.
- Los bienes públicos globales.

La lista no es estática, sino que cambia como respuesta a los eventos globales.

Referencias y lecturas adicionales

Adams, Richard H. Jr. y John Page. 2003. "International Migration, Remittances and Poverty in Developing Countries". Documento de trabajo 3179 sobre investigación de políticas del Banco Mundial, Washington, DC.

Banco Mundial. s.f. "About Private Participation in Infrastructure". http://www.worldbank.org/infrastructure/ppi/.

———. 1994. *Enhancing Women's Participation in Economic Development*. Washington, DC: Banco Mundial. http://books.google.ca/books?id=CJBmEClPci8C&dq=World+Bank.+1994.+Enhancing+Women%E2%80%99s+Participation+in+Economic+Development.+Washington,+DC:+World+Bank.&printsec=frontcover&source=bn&hl=en&ei=IPy2SdntFYueMoTU_NcK&sa=X&oi=book_result&resnum=4&ct=result#PPA5,M1.

———. 1996b. *World Bank Participation Sourcebook*. http://www.worldbank.org/wbi/sourcebook/sbhome.htm.

———. 1999. "Monitoring and Evaluation Capacity Development in Africa". *Précis* Spring (183). http://wbln0018.worldbank.org/oed/oeddoclib.nsf/7f2a291f9f1204c685256808006a0025/34b9bade34aca617852567fc00576017/US$FILE/183precis.pdf.

———. 2001. *Strategic Directions for FY02-FY04.* Washington, DC. http://lnweb18.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/762997a38851fa0685256f8200777e15/US$FILE/gppp_main_report_phase_2.pdf#page=21.

———. 2003. *Global Development Finance 2003.* Washington, DC: Banco Mundial. http://siteresources.worldbank.org/INTRGDF/Resources/GDF0slide0-show103010DC0press0launch.pdf.

———. 2004a. "Evaluating the World Bank's Approach to Global Programs: Addressing the Challenges of Globalization". Independent Evaluation Group, Washington, DC. http://www.worldbank.org/oed/gppp/.

———. 2004b. *Informe sobre el Seguimiento Mundial 2004: Políticas y acciones para ampliar los ODM*. Washington, DC: Banco Mundial.

———. 2005a. *Informe sobre el Seguimiento Mundial 2005: Educación para todos.* Washington, DC: Banco Mundial.

———. 2005b. *Informe sobre el Desarrollo Mundial 2005: un mejor clima de inversión para todos.* Washington, DC: Banco Mundial.

———. 2006a. *Informe sobre el Seguimiento Mundial 2006: Equidad y desarrollo.* Washington, DC: Banco Mundial.

———. 2007a. *Aid Architecture: An Overview of the Main Trends in Official Development Assistance Flows, Executive Summary.* http://siteresources.worldbank.org/IDA/Resources/Aidarchitecture-execsummary.pdf.

———. 2007b. *Conflict Prevention and Reconstruction.* http://lnweb18.worldbank.org/ESSD/sdvext.nsf/67ByDocName/ConflictPreventionandReconstruction.

———. 2007c. *The Data Revolution: Measuring Governance and Corruption.* http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:20190210~menuPK:34457~pagePK:34370~piPK:34424~theSitePK:4607,00.html.

———. 2007d. *Doing Business: Economy Profile Reports*. Washington, DC: Banco Mundial. http://rru.worldbank.org/DoingBusiness/.

———. 2007e. *The Enhanced HIPC Initiative: Overview.* http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTDEBTDEPT/0,contentMDK:21254881~menuPK:64166739~pagePK:64166689~piPK:64166646~theSitePK:469043,00.html.

———. 2007f. *Informe sobre el Seguimiento Mundial 2007: Development and the Next Generation.* Washington, DC: Banco Mundial.

———. 2008. *Online Atlas of the Millennium Development Goals: Building a Better World*. http://devdata.worldbank.org/atlas-mdg/.

BankTrack. 2008. *The Equator Principles*. http://www.banktrack.org/.

Barslund, Mikkel y Finn Tarp. 2007. "Formal and Informal Rural Credit in Tour Provinces of Vietnam". Documento de discusión 07-07, Department of Economics, University of Copenhague.

BEER (U.K. Department for Business Enterprise and Regulatory Reform). 2004. "What Is CSR?" Londres. http://www.csr.gov.uk/whatiscsr.shtml.

Boyle, Richard. 2002. "A Two-Tiered Approach: Evaluation Practice in the Republic of Ireland". En *International Atlas of Evaluation*, eds. Jan-Eric Furubo, Ray Rist y Rolf Sandahl, 261-72. Nueva Brunswick, NJ: Transaction Publishers.

———. 2005. *Evaluation Capacity Development in the Republic of Ireland*. ECD, Serie N° 14 de documentos de trabajo, Banco Mundial, Evaluation Capacity Development, Washington, DC.

Camdessus, Michael. 1998. "Money Laundering: The Importance of International Countermeasures". Discurso del Director Administrativo del Fondo Monetario Internacional, Washington, DC.

CFI (Corporación Financiera Internacional). 2004. *Strengthening the Foundations for Growth and Private Sector Development: Investment Climate and Infrastructure Development*. Development Committee, Washington, DC. http://siteresources.worldbank.org/IDA/Resources/PSDWBGEXT.pdf.

———. 2007. *Business Environment Snapshots*. Washington, DC. http://rru.worldbank.org/documents/BES_Methodology_Note_External.pdf.

CGAP (Consultative Group to Assist the Poor). 2003. *CGAP Publications on Assessment and Evaluation*. http://www.cgap.org/portal/site/CGAP/menuitem.9fab704d4469eb0167808010591010a0/.

ChannahSorah, Vijaya Vinita. 2003. "Moving from Measuring Processes to Outcomes: Lessons Learned from GPRA in the United States". Documento presentado en la conferencia conjunta del Banco Mundial y el Korea Development Institute sobre "Performance Evaluation System and Guidelines with Application to Large-Scale Construction, RyD y Job Training Investments", Seúl, julio 24-25.

Chelimsky, Eleanor y William R. Shadish, eds. 1997. *Evaluation for the 21st Century: A Handbook*. Thousand Oaks, CA: Sage Publications.

Chemin, Matthieu. 2008. "Special Section on Microfinance. The Benefits and Costs of Microfinance: Evidence from Bangladesh". *Journal of Development Studies* 44 (4): 463-84.

Collier, Paul, V., L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol y Nicholas Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Washington, DC: Oxford University Press para el Banco Mundial.

Economic Commission for Europe. 1998. *Public-Private Partnerships: A New Concept for Infrastructure Development.* BOT Expert Group, Naciones Unidas, Nueva York. http://rru.worldbank.org/Documents/Toolkits/Highways/pdf/42.pdf.Equator Principles. 2004. http://www.fatf-gafi.org/dataoecd/14/53/38336949.pdf.

———. 2007. "A Milestone or Just Good PR?" http://www.equator-principles.com/principles.shtml.

Feuerstein, M. T. 1986. *Partners in Evaluation: Evaluating Development and Community Programs with Participants*. Londres: MacMillan, en asociación con Teaching Aids at Low Cost.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines*. Nueva York: Pearson Education, Inc.

FMI (Fondo Monetario Internacional). 2009. "A Fact Sheet: Debt Relief under the Heavily Indebted Poor Countries (HIPC) Initiative". http://www.info.org/external/np/exr/facts/hipc.htm.

Friis, H. 1965. *Development of Social Research in Ireland*. Institute of Public Administration, Dublín.

Fukuda-Parr, Sakiko, Carlos Lopes y Khalid Malik, eds. 2002. *Capacity for Development: New Solutions to Old Problems*. Londres: Earthscan Publications.

Furubo, Jan-Eric y Rolf Sandahl. 2002. "Coordinated Pluralism". En *International Atlas of Evaluation*, eds. Jan-Eric Furubo, Ray Rist y Rolf Sandahl, 115-28. Nueva Brunswick, NJ: Transaction Publishers.

Furubo, Jan-Eric, Ray Rist y Rolf Sandahl, eds. 2002. *International Atlas of Evaluation.* Nueva Brunswick, NJ: Transaction Publishers.

GAFI (Grupo de Acción Financiera contra el Blanqueo de Capitales). 2007. *Monitoring the Implementation of the Forty Recommendations*. http://www.fatf-gafi.org/document/60/0,3343, en_32250379_32236920_34039228_1_1_1_1,00.html.

G-8. 2004a. *G-8 Action Plan: Applying the Power of Entrepreneurship to the Eradication of Poverty*. Cumbre de Sea Island, junio.

———. 2004b. *G-8 Plan of Support for Reform*. Cumbre de Sea, junio.

Gerrard, Christopher. 2006. "Global Partnership Programs: Addressing the Challenge of Evaluation". Presentación en PowerPoint. http://www.oecd.org/secure/pptDocument/0,2835, en_21571361_34047972_36368404_1_1_1_1,00.ppt.

gpgNet. 2008. *The Global Network on Global Public Goods*. http://www.sdnp.undp.org/gpgn/#.

Graham, Jeffrey P. y R. Barry Spaulding. 2005. *Going Global: Understanding Foreign Direct Investment.* JPG Consulting. http://www.going-global.com/articles/understanding_foreign_direct_investment.htm.

Guerrero, R. Pablo. 1999. "Evaluation Capacity Development: Comparative Insights from Colombia, China e Indonesia". En *Building Effective Evaluation Capacity: Lessons from Practice*, eds. Richard Boyle y Donald Lemaire, 177-94. New Brunswick, NJ: Transaction Publishers.

Hatry, Harry P., Elaine Morely, Shelli B. Rossman y Joseph P. Wholey. 2003. *How Federal Programs Use Outcome Information: Opportunities for Federal Managers*. IBM Endowment for the Business of Government, Washington, DC.

Hauge, Arild. 2001. "Strengthening Capacity for Monitoring and Evaluation in Uganda: A Results-Based Perspective". ECD, Documento de trabajo 8, Banco Mundial, Operations Evaluation Department, Evaluation Capacity Development, Washington, DC.

Hougi, Hong y Ray C. Rist. 2002. "Evaluation Capacity Building in the People's Republic of China". En *International Atlas of Evaluation*, eds. Jan-Eric Furubo, Ray Rist y Rolf Sandahl, 249-60. Nueva Brunswick, NJ: Transaction Publishers.

IDA (International Development Association) y FMI (Fondo Monetario Internacional). 2007. *Heavily Indebted Poor Countries HIPC Initiative and Multilateral Debt Relief Initiative MDRI: Status of Implementation.* http://siteresources.worldbank.org/DEVCOMMINT/Documentation/21510683/ DC2007-0021 E HIPC.pdf.

IDS (Institute of Development Studies). 2008. "Impact Evaluation: The Experience of Official Agencies". *IDS Bulletin* 39 (1). http://www.ntd.co.uk/idsbookshop/details.asp?id=1030.

IEG (Independent Evaluation Group). 2008. "Doing Business: An Independent Evaluation. Taking the Measure of the World Bank-IFC Doing Business Indicators".

IFAC (International Federation of Accountants). 2004. *Anti-Money Laundering,* 2ª ed. Nueva York. 2ª

Investor Dictionary.com. 2006. "Money Laundering". http://www.investordictionary.com/definition/ money+laundering.aspx.

Jalan, Bimal. 2000. "Reflections on Gender Policy". En *Evaluating Gender Impact of Bank Assistance*, 75-76. Banco Mundial, Operations Evaluation Department, Washington, DC.

Joint Progress toward Enhanced Aid Effectiveness, High-Level Forum. 2005. *Paris Declaration on Aid Effectiveness: Ownership, Harmonization, Alignment, Results, and Mutual Accountability.* http://www1.worldbank.org/harmonization/Paris/FINALPARISDECLARATION.pdf/.

Katjomulse, Kavazeua, Patrick N. Osakwe, Abebe Shimeles y Sher Verick. 2007. *The Monterrey Consensus and Development in Africa: Progress, Challenges, and Way Forward*. United Nations Economic Commission for Africa (UNECA), Addis Abeba.

Kreimer, Alcira, John Eriksson, Robert Muscat, Margaret Arnold y Colin Scott. 1998. *The World Bank's Experience with Post-Conflict Reconstruction.* Banco Mundial, Operations Evaluation Department, Washington, DC.

http://lnweb90.worldbank.org/oed/oeddoclib.nsf/b57456d58aba40e585256ad400736404/f753e43e728a27b38525681700503796/$FILE/PostCon.pdf.

Kusek, Jody Zall, Ray C. Rist y Elizabeth M. White. 2004. "How Will We Kong Millennium Development Results When We See Them? Building a Results-Based Monitoring and Evaluation System to Give Us the Answer". Banco Mundial, Documento de trabajo 66, Región África, Washington, DC.

Lawrence, J. 1989. "Engaging Recipients in Development Evaluation: The 'Stakeholder' Approach". *Evaluation Review* 13 (3): 243-56.

Lee, Yoon-Shik. 1999. "Evaluation Coverage". En *Building Effective Evaluation Capacity: Lessons from Practice*, eds. Richard Boyle y Donald Lemaire, 75-91. Nueva Brunswick, NJ: Transaction Publications.

Linux Information Project. 2006. "Public Goods: A Brief Introduction". http://www.linfo.org/public_good.html.

Mackay, Keith. 2002. "The Australian Government: Success with a Central, Directive Approach". En *International Atlas of Evaluation*, eds. Jan-Eric Furubo, Ray C. Rist y Rolf Sandahl, 157-74. Nueva Brunswick, NJ: Transaction Publishers.

———. 2007. "Three Generations of National M&E Systems in Australia". Presentación en PowerPoint para la Third Latin America and Caribbean Regional Conference on Monitoring and Evaluation, Lima, julio 23-24.

———. 2008. *M&E Systems to Improve Government Performance: Lessons from Australia, Chile and Colombia*. Presentación en PowerPoint para la Delegación de Alto Nivel de la República Popular de China, Washington, DC, marzo 6.

Moser, Caroline O. N. 1995. "Evaluating Gender Impacts". *New Directions for Evaluation* 67 (otoño): 105-17.

NCSTE (China National Center for Evaluation of Science and Technology) e IOB (Netherland's Policy and Operations Evaluation Department). 2004. *A Country-Led Joint Evaluation of the ORET/MILIEV Programme in China.* http://www.euforic.org/iob/docs/200610201336433964.pdf.

Nellis, John. 1999. "Time to Rethink Privatization in Transition Economies?" IFC, Documento de discusión 38, International Finance Corporation, Washington, DC.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2002a. *OECD Glossary of Statistical Terms*. http://stats.oecd.org/glossary/index.htm.

———. 2002b. "Public Management and Governance: Overview of Results-Focused Management and Budgeting in OECD Member Countries". Documento presentado en la vigésima tercera reunión anual de funcionarios superiores de presupuesto, Washington, DC, junio 3-4.

———. 2003. *Joint OECD DAC/Development Centre Experts' Seminar on Aid Effectiveness and Selectivity: Integrating Multiple Objectives into Aid Allocations*. Development Assistance Committee. http://www.oecd.org/document/51/0,2340,en_2649_34435_2501555_119808_1_1_1,00.html.

———. 2004a. *Final ODA Data for 2003*. http://www.oecd.org/dataoecd/19/52/34352584.pdf.

———. 2004b. *The Istanbul Ministerial Declaration on Fostering Growth of Innovative and Internationally Competitive SME*s. http://www.oecd.org/document/16/0,3343,en_2649_201185_32020176_1_1_1_1,00.html.

———. 2005a. "Aid Rising Sharply, According to Latest OECD Figures". http://www.oecd.org/dataoecd/0/41/35842562.pdf.

———. 2005b. *The Paris Declaration*. Development Co-operation Directorate. http://www.oecd.org/document/18/0,2340,en_2649_3236398_35401554_1_1_1_1,00.html.

———. 2006. *Final ODA Data for 2005.* http://www.oecd.org/dataoecd/52/18/37790990.pdf.

———. 2007a. "Development Aid from OECD Countries Fell 5.1 Percent in 2006". http://www.oecd.org/document/17/0,3343,en_2649_33721_38341265_1_1_1_1,00.html.

———. 2007b. *Final ODA Flows in 2006*. http://www.oecd.org/dataoecd/7/20/39768315.pdf.

———. 2007c. "Monitoring the Paris Declaration". http://www.oecd.org/department/0,3355, en_2649_15577209_1_1_1_1_1,00.html.

OCDE y PUMA (Public Management Committee). 2002. "Overview of Results-Focused Management and Budgeting in OECD Member Countries". Vigésima tercera reunión anual de funcionarios superiores de presupuesto de la OCDE, Washington, DC, junio 3-4.

Oxford Analytica. 2004a. "Foundations Muscle into Aid Arena". Agosto 10, Oxford.

———. 2004b. "Remittances Fund Investment Growth". Septiembre 7, Oxford.

Patton, M. Q. 1997. *Utilization-Focused Evaluation*. 3ª ed. Thousand Oaks, CA: Sage Publications.

———. 2006. "Recent Trends in Evaluation". Documento presentado a la Corporación Financiera Internacional, Washington, DC, mayo 8.

Picciotto, Robert. 2002a. "Development Cooperation and Performance Evaluation: The Monterrey Challenge". Banco Mundial, Operations Development Department, Washington, DC.

———. 2002b. "Development Evaluation as a Discipline". Presentación del International Program for Development Evaluation Training (IPDET), Ottawa, junio.

———. 2003. "International Trends and Development Evaluation: The Need for Ideas". *American Journal of Evaluation* 24: 227-34.

Picciotto, Robert y Ray C. Rist. 1995. *Evaluating Country Development Policies and Programs: New Approaches and a New Agenda.* San-Francisco: Jossey-Bass Publishers.

Pollin, Robert. 2007. *Microcredit: False Hopes and Real Possibilities.* Foreign Policy in Focus. http:// www.fpif.org/fpiftxt/4323.

Public Services Organisation Review Group. 1969. *Report of the Public Services Organization Review Group.* Dublín: Stationary Office.

Qureshi, Zia. 2004. *Millennium Development Goals and Monterrey Consensus: From Vision to Action.* Banco Mundial, Washington, DC. http://wbln0018.worldbank.org/eurvp/web.nsf/Pages/ Paper+by+Qureshi/US$File/MOHAMMED+QURESHI.PDF.

Rasappan, Arunaselam. 2007. "Implementation Strategies and Lessons Learnt with Results-Based Budgeting Malaysia". Curso de capacitación sobre presupuestos de programa y desempeño, ITP Pune, India, octubre 1-5. http:// blog-pfm.imf.org/pfmblog/files/rasappan_implementation_strategies_lessons_malaysia.pdf.

República de Francia. 2001. Ministerio de Economía, Finanzas e Industria. "Towards New Public Management". *Newsletter on the Public Finance Reform* 1 (septiembre), París.

Rist, Ray C. y Nicoletta Stame, eds. 2006. *From Studies to Streams: Managing Evaluative Systems.* Nueva Brunswick, NJ: Transaction Books.

Schacter, Mark. 2000. "Sub-Saharan Africa: Lessons from Experience in Supporting Sound Governance". ECD, Documento de trabajo 7, Banco Mundial, Evaluation Capacity Department, Washington, DC.

Soros Foundations Network. 2007. "About OSI and the Soros Foundation Network". http://www.soros.org/about/overview.

Takamasa, Akiyama y Kondo Masanori, eds. 2003. "Global ODA since the Monterrey Conference". Foundation for Advanced Studies on International Development (FASID), International Development Research Institute, Tokio. http://www.fasid.or.jp/english/publication/research/pdf/ global.pdf.

Tavistock Institute, en asociación con GHK e IRS. 2003. *The Evaluation of Socio-Economic Development: The GUIDE.* http://coursenligne.sciences-po.fr/2004_2005/g_martin/guide2.pdf.

Thomas, Koshy. 2007. "Integrated Results-Based Management in Malaysia". En *Results Matter: Ideas and Experiences on Managing for Development Results.* Banco Asiático de Desarrollo. http://www. adb.org/Documents/Periodicals/MfDR/dec-2007.pdf.

Trosa, Sylvie. 2008. "Towards a Postbureaucratic Management in France". *Politique et management public* (2).

Uganda Office of the Prime Minister (OPM). 2007a. *National Integrated Monitoring and Evaluation Strategy (NIMES): 2006-2007 Bi-Annual Implementation Progress Report.* Kampala.

———. 2007b. "Working Note: Monitoring and Evaluation of the National Development Plan". Octubre, Kampala.

Unctad (Conferencia de las Naciones Unidas sobre Comercio y Desarrollo). 2008. "Foreign Direct Investment Reached New Record in 2007". Comunicado de prensa. http://www.unctad.org/Templates/Webflyer.asp?docID=9439&intItemID=2068&lang=1.

Uneca (United Nations Economic Commission for Africa). 2007. "Financing Development Section, Trade, Finance y Economic Commission for Africa, Addis Abeba". http://www.uneca.org/eca_programmes/trade_and_regional_integration/documents/MonterreyConsensusMainReport.pdf.

United Nations Office on Drugs and Crime. 2006. *The United Nations Convention against Transnational Organized Crime and Its Protocols*. http://www.unodc.org/unodc/en/treaties/CTOC/index.html.

U.S. Department of Labor. 1993. Government Performance and Results Act. Employment and Training Administration, Washington, DC. http://www.doleta.gov/performance/goals/gpra.cfm.

U.S. GAO (Government Accountability Office). 2003. *Executive Guide: Effectively Implementing the Government Performance and Results Act.* http://www.gao.gov/special.pubs/gpra.htm.

Woroniuk, B. y J. Schalwyk. 1998. "Donor Practices: Evaluation and Preparation Tipsheet". OCDE, París.

Sitios web

Finanzas

Banco Mundial. *Doing Business: Economy Profile Reports*. http://rru.worldbank.org/DoingBusiness/.

CGAP (Consultative Group to Assist the Poor). http://www.cgap.org/.

———. *Assessment and Evaluation*. http://www.cgap.org/publications/assessment_evaluation.html.

The Equator Principles. http://www.equator-principles.com/index.shtml.

Tedeschi, Gwendolyn Alexander. 2008. *Overcoming Selection Bias in Microcredit Impact Assessments: A Case Study in Peru.* http://www.informaworld.com/smpp/content~content=a792696580~db=all~order=page.

Género

OCDE (Organización para la Cooperación y el Desarrollo Económicos). "Gender Tipsheet, Evaluation". Development Assistance Committee. http://www.oecd.org/dataoecd/2/13/1896352.pdf.

Gobernanza

Banco Mundial. 2007a. *Governance Matters*. http://info.worldbank.org/governance/wgi2007/ and http://info.worldbank.org/governance/wgi2007/pdf/booklet_decade_of_measuring_governance.pdf.

———. 2007b. video sin título. http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:21400275~menuPK:51416191~pagePK:64257043~piPK:437376~theSitePK:4607,00.html.

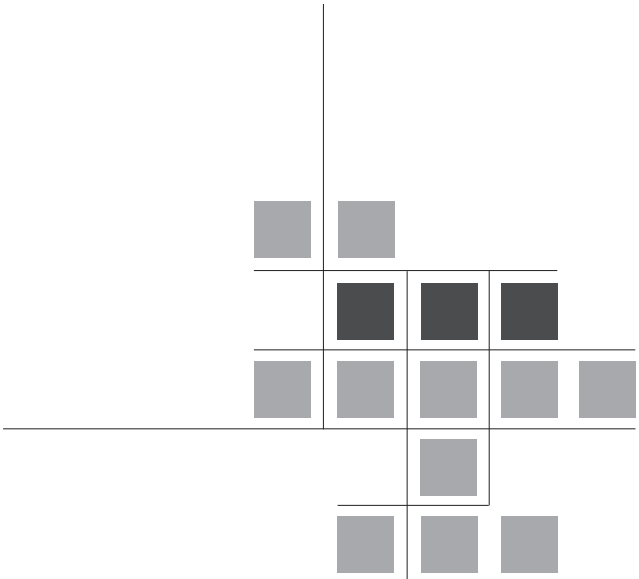
Transparency International. http://www.transparency.org/.

Objetivos de Desarrollo del Milenio

Naciones Unidas. *Millennium Development Goals*. http://www.un.org/millenniumgoals/.

Pobreza

Banco Mundial. Software AdePT para facilitar y efectuar más rápidamente el análisis de la pobreza. econ.worldbank.org/programs/poverty/adept.
———. PovertyNet newsletter. http://www.worldbank.org/poverty.
Poverty-Environment Web site. http://www.povertyenvironment.net.



PREPARAR Y REALIZAR EVALUACIONES EFICACES PARA EL DESARROLLO

"Nuestros planes se malogran porque carecen de propósito. Cuando un hombre no sabe hacia qué puerto se dirige, ningún viento es correcto".

—SÉNECA

Capítulo 3. Construir un sistema de seguimiento y evaluación basado en resultados

- Importancia del control y la evaluación basados en resultados
- ¿Qué son seguimiento y evaluación basados en resultados?
- Seguimiento y evaluación tradicionales frente a basados en resultados
- Diez pasos para construir un sistema de seguimiento y evaluación basado en resultados

Capítulo 4. Conocimiento del contexto de evaluación y Teoría Programática del Cambio

- Análisis *Front-End*
- Identificación del cliente principal y de las partes interesadas clave
- Comprensión del contexto
- Investigación del conocimiento existente
- Construir, utilizar y evaluar una Teoría del Cambio

Capítulo 5. Consideración sobre el enfoque de la evaluación

- Enfoques generales en la evaluación
- Desafíos en el avance

001El camino.indd 93 08/11/2010 03:27:44 p.m.

CAPÍTULO 3

Construir un sistema de seguimiento y evaluación basado en resultados

En todo el mundo, los gobiernos están intentando responder a las demandas y presiones para mejorar la vida de los ciudadanos. Las presiones y demandas internas sobre los gobiernos y las organizaciones de desarrollo son la causa de que estén a la búsqueda de nuevas formas de mejorar la administración pública. Las mejoras pueden incluir una mayor rendición de cuentas y transparencia, y una mejor eficacia de las intervenciones. El seguimiento y evaluación (SyE) basados en resultados son una herramienta de gestión para hacer seguimiento al progreso y demostrar el impacto de los proyectos, programas y políticas para el desarrollo.

Importancia del seguimiento y la evaluación basados en resultados

Cada vez más existen presiones en los países en desarrollo para mejorar el desempeño de los sectores públicos, y la respuesta a estas presiones lleva a los países a desarrollar sistemas de administración del desempeño. Estos nuevos sistemas implican reformas en presupuestos, recursos humanos y cultura organizativa. Para evaluar si los esfuerzos del sector público están funcionando, también es necesario medir el desempeño. Los

001El camino.indd 95 08/11/2010 03:27:44 p.m.

sistemas de SyE siguen la pista a los resultados producidos (o no producidos) por los gobiernos y otras entidades.

Muchas iniciativas están impulsando a los gobiernos a adoptar sistemas de administración pública que muestren resultados. Los Objetivos de Desarrollo del Milenio (ODM) y la Iniciativa PPAE constituyen dos ejemplos de estas iniciativas.

La estrategia esbozada en este capítulo se basa en la experiencia de los países desarrollados, especialmente los de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), pero refleja también los retos y las dificultades que se presentan en muchos países en desarrollo cuando éstos intentan iniciar sistemas de medición del desempeño. Los problemas de estos países van desde la falta de capacidad de destrezas hasta deficientes estructuras de gobernabilidad y sistemas que están muy lejos de ser transparentes. Aunque el enfoque primario de este capítulo sea mejorar la eficacia del gobierno y la rendición de cuentas, utilizando un sistema sólido de seguimiento y evaluación (SyE), los principios y estrategias se aplican igualmente a las organizaciones, políticas, programas y proyectos.

Un sistema de SyE basado en resultados suministra información crucial sobre el sector público o el desempeño organizativo, y puede ayudar a los autores de políticas, a quienes toman las decisiones y a otras partes interesadas a responder las preguntas fundamentales sobre si se cumplieron las promesas y se obtuvieron los *efectos directos.* El SyE es el medio a través del cual pueden demostrarse las mejoras, o la falta de éstas (recuadro 3.1).

Efecto directo: Los beneficios que se obtienen de un proyecto, programa o política (un resultado implica cambio conductual u organizativo y no puede comprarse).

Al informar sobre los resultados de distintas intervenciones, los gobiernos y otras organizaciones promueven la credibilidad y la confianza pública en su labor. El suministro de información al público apoya también la agenda para el desarrollo, la cual se está desplazando hacia una mayor rendición de cuentas sobre los préstamos de la ayuda.

Un buen sistema de SyE basado en resultados puede ser extremamente útil como herramienta administrativa y motivacional, y contribuye a centrar la atención en el logro de efectos directos importantes para la organización y sus partes interesadas, además de que aporta un motivo para establecer metas y objetivos clave que se desprendan de estos resultados.

También suministra a los administradores información crucial sobre si la estrategia que orienta la intervención es apropiada, correcta y adecuada a los cambios que se pretenden mediante la intervención.

Un buen sistema de SyE constituye también una fuente esencial de información para modernizar y mejorar las intervenciones, a fin de maximizar la probabilidad de éxito. Contribuye a identificar las intervenciones promisorias en forma temprana de manera que puedan implementarse potencialmente en otras partes. Contar con datos disponibles sobre lo bien que funciona un proyecto, programa o política particular, aporta información útil para formular y justificar solicitudes de presupuesto, y permite

Recuadro 3.1 *El poder de medir los resultados*

La medición de resultados es de importancia crítica por las razones siguientes:

- Si no se miden los resultados, no puede distinguirse el éxito del fracaso.
- Si no puede identificarse el éxito, no puede recompensarse.
- Si no se recompensa el éxito, probablemente se está recompensando el fracaso.
- Si no se puede identificar el éxito, no puede aprenderse con él.
- Si no puede reconocerse el fracaso, no puede corregirse.
- Si pueden demostrarse efectos directos, puede obtenerse el apoyo público.

Fuente: Osborn y Gaebler, 1992.

también la asignación juiciosa de los recursos escasos a las intervenciones que ofrecen los mayores beneficios.

Una vez establecidos los efectos directos, seleccionados los indicadores y fijados los objetivos, la organización puede esforzarse para lograrlos. Un sistema de SyE puede suministrar información oportuna al personal sobre el progreso, y contribuir a identificar temprano las debilidades que requieran acción correctiva. El seguimiento de los datos suministra igualmente información sobre casos extremos (sitios con un desempeño particularmente bueno o particularmente deficiente), y la evaluación puede efectuarse para averiguar qué explica los casos extremos.

Seguimiento basado en resultados: Proceso continuo de recopilar y analizar información sobre indicadores clave con el fin de medir el progreso en el cumplimiento de los objetivos.

Objetivo: Una cantidad cuantificable de cambio que debe obtenerse en un tiempo especificado en un indicador.

Indicador: Medida controlada en forma sistemática a través del tiempo, que indica el progreso (o la falta de progreso) hacia el cumplimiento de un objetivo.

Evaluación basada en resultados: Evaluación de una intervención planeada, en curso, o terminada, con el fin de establecer su pertinencia, eficiencia, eficacia, impacto y sostenibilidad.

Impacto: Un beneficio (resultado) a largo plazo que se obtiene cuando se logra una serie de resultados (los Objetivos de Desarrollo del Milenio constituyen declaraciones de impacto).

¿Qué son seguimiento y evaluación basados en resultados?

La información basada en resultados puede provenir de dos fuentes complementarias: un sistema de seguimiento y un sistema de evaluación (recuadro 3.2). Ambos sistemas son esenciales para una medición efectiva del desempeño.

Seguimiento y evaluación tradicionales frente a los basados en resultados

Los gobiernos han empleado desde hace mucho tiempo SyE tradicional, por ejemplo, con el control de sus gastos y rentas, niveles de personal y recursos, actividades de programas y proyectos, número de participantes y producción de bienes y servicios. Sin embargo, debe distinguirse entre el SyE tradicional y el basado en resultados:

Recuadro 3.2 Diferencia entre seguimiento basado en resultados y evaluación basada en resultados

El *seguimiento basado en resultados* es el proceso continuo de recopilar y analizar información sobre indicadores clave, y comparar los resultados reales con los esperados a fin de medir qué tan bien se está ejecutando un proyecto, programa o política. Es un proceso continuo de medir el progreso hacia la obtención de resultados explícitos a corto, mediano y largo plazo, siguiendo la evidencia del movimiento hacia el logro de *objetivos* específicos predeterminados mediante el uso de *indicadores*. El seguimiento basado en resultados provee retroalimentación sobre el progreso (o la falta de progreso) al personal y a quienes toman las decisiones, quienes pueden utilizar la información en diversas formas para mejorar el desempeño.

La *evaluación basada en resultados* es una valorización de una intervención planeada, en curso o terminada, con el fin de establecer su pertinencia, eficiencia, eficacia, *impacto* y sostenibilidad. La intención es suministrar información verosímil y útil, haciendo posible que las enseñanzas aprendidas se incorporen al proceso de toma de decisiones de los receptores. La evaluación adopta un punto de vista más amplio de una intervención, cuestionando si el progreso hacia el objetivo o el resultado explícito es causado por la intervención o si existe alguna otra explicación para los cambios recogidos por el sistema de control. Las preguntas de la evaluación pueden incluir las siguientes:

- ¿Son pertinentes los objetivos y efectos directos?
- ¿Con qué eficiencia y eficacia se están obteniendo?
- ¿Qué efectos imprevistos ha causado la intervención?
- ¿Representa la intervención la estrategia más efectiva en costos y sostenible para satisfacer un conjunto particular de necesidades?

- El *SyE tradicional* se centra en el control y la evaluación de los insumos, actividades y productos (es decir, en la ejecución de los proyectos y programas).
- El *SyE basado en resultados* combina el enfoque tradicional de controlar la ejecución con la evaluación de los efectos directos e impactos, o más generalmente de los resultados.

SyE tradicionales: Centrados en la ejecución de proyectos o programas.

SyE basados en resultados: Combinan el enfoque tradicional con la evaluación de los resultados.

Teoría del cambio: Hipótesis sobre cómo una iniciativa conduce a los resultados deseados.

Esta vinculación del progreso en la ejecución con el progreso en el logro de los objetivos o los resultados deseados de las políticas y programas del gobierno, es lo que hace tan útiles al SyE basados en resultados como herramienta de administración pública. La implementación de esta clase de sistema de SyE permite a la organización modificar y efectuar ajustes a la teoría del cambio y a los procesos de ejecución con el fin de apoyar más directamente el logro de los objetivos y efectos directos deseados.

La teoría del cambio

Una forma de ver las diferencias entre el SyE tradicionales y los basados en resultados es considerar la teoría del cambio. De acuerdo con Kusek y Rist (2004), la teoría del cambio es una representación de cómo se espera que una intervención conduzca a los resultados deseados. (Mayor información sobre la teoría del cambio y definiciones se suministra en el capítulo 4.) Los modelos de la teoría del cambio suelen tener cinco componentes: insumos, actividades, productos, efectos directos e impactos (cuadro 3.1). Algunos modelos de la teoría del cambio incluyen también otras características, como grupos objetivo y factores internos y externos.

Una teoría del cambio puede presentarse en forma gráfica (gráfico 3.1), y este modelo puede utilizarse para enmarcar un enfoque basado en resultados, hacia un problema como reducir la morbilidad infantil con terapia de rehidratación oral (gráfico 3.2).

Cuadro 3.1 *Principales componentes de una teoría del cambio*

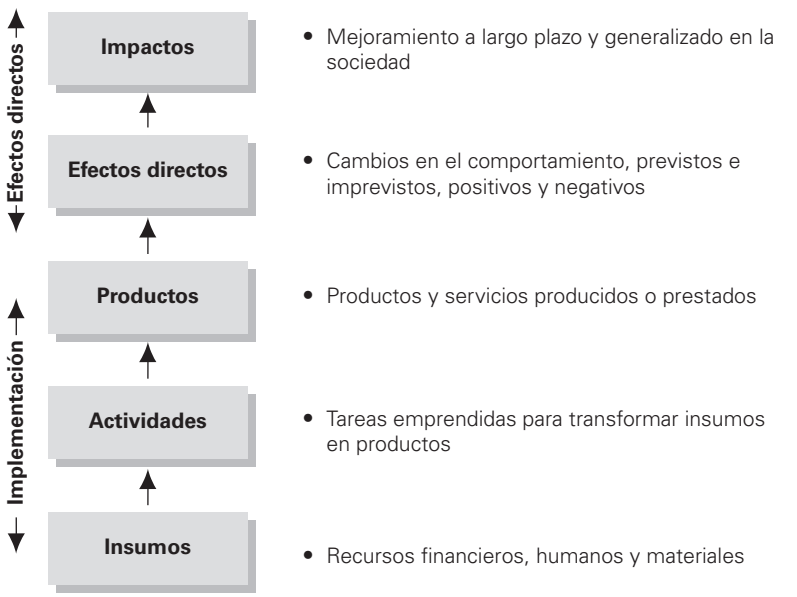
Componente	*Descripción*
Insumos	Recursos que van a un proyecto, programa o política (fondos, personal, equipo, materiales de plan de estudios, etcétera).
Actividades	Lo que se hace. Las actividades pueden expresarse con un verbo ("mercadear", "proveer", "facilitar", "entregar").
Productos	Lo que se produce. Son los productos tangibles o los servicios que se producen con las actividades, y usualmente se expresan con sustantivos. De manera habitual no tienen modificadores, son tangibles y pueden contarse.
Resultados	Por qué se hace. Los efectos directos son los cambios de comportamiento que resultan de los productos del proyecto (dejar de fumar, hervir el agua, utilizar redes hospitalarias). Los efectos directos se pueden aumentar, disminuir, mejorar o mantener.
Impactos	Los cambios a largo plazo resultantes de una acumulación de los efectos directos. Pueden ser similares a los objetivos estratégicos.

Fuente: Kusek y Rist, 2004.

Indicadores de desempeño

El control incluye la medición del progreso para obtener un efecto directo o un impacto (resultados). Sin embargo, el efecto directo no puede medirse en forma directa, sino que debe primero representarse con un conjunto de indicadores que, al medirse regularmente, suministran información sobre si se están obteniendo, o no, los efectos directos o impactos. Un indicador de desempeño es "una variable que permite verificar los cambios en la intervención para el desarrollo, o muestra resultados relacionados con lo planeado" (OCDE, 2002, p. 29).

Gráfico 3.1 *Teoría programática del cambio (modelo lógico) para obtener efectos directos e impactos*



Fuente: Binnendijk, 2000.

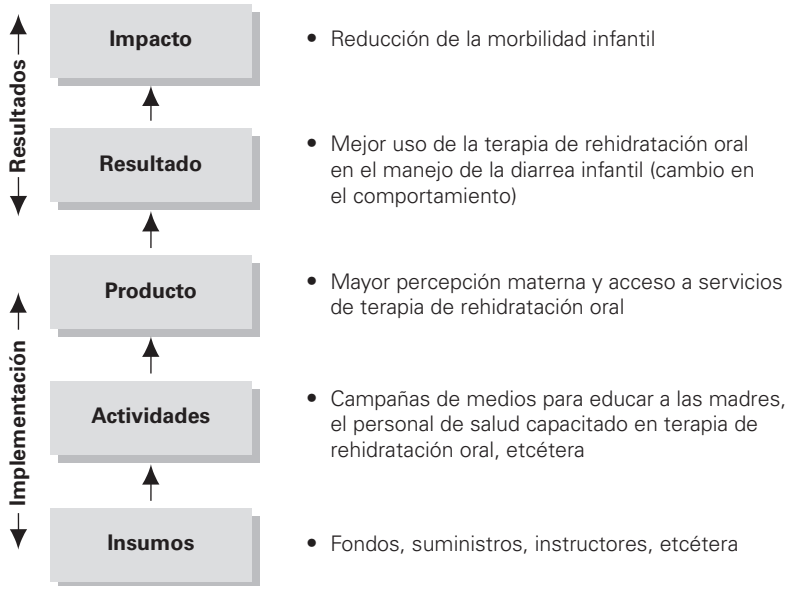
Por ejemplo, si un país selecciona el objetivo de mejorar la salud de los niños, reduciendo la morbilidad infantil de enfermedades infecciosas en un 30% en los próximos cinco años, debe primero identificar un conjunto de indicadores que representen los cambios en la incidencia de la morbilidad infantil de enfermedades infecciosas en medidas más específicas. Los indicadores que pueden ayudar a evaluar los cambios en la morbilidad infantil pueden incluir los siguientes:

- La incidencia y la prevalencia de enfermedades infecciosas, como la hepatitis (un determinante directo).
- El nivel de salud materna (un determinante indirecto).
- El grado al cual los niños tienen acceso a agua limpia.

La evidencia acumulada de un grupo de indicadores es lo que los administradores examinan para comprobar si el programa está progresando. Ningún efecto directo o impacto debería medirse con un solo indicador.

La medición de un conjunto desagregado de indicadores (un conjunto de indicadores que ha sido dividido en sus partes constitutivas) suministra información importante en cuanto hace al buen o mal funcionamiento de los programas y políticas oficiales, para

Gráfico 3.2 *Ejemplo de teoría programática del cambio (modelo lógico) para reducir la morbilidad infantil mediante el uso de terapia de rehidratación oral*



Fuente: Adaptado de Binnendijk, 2000.

obtener el efecto directo o el impacto deseado. También se utilizan para identificar sitios de mejor o peor desempeño que el promedio (casos extremos del programa) y políticas que tienen o no un buen desempeño. Si, por ejemplo, el conjunto de indicadores revela que, a través del tiempo, cada vez menos niños tienen acceso a agua limpia, el gobierno puede utilizar esta información para reformar los programas dirigidos a mejorar el suministro de agua, o fortalecer los programas que suministran información a los padres sobre la necesidad de desinfectar el agua antes de proveérsela a los niños.

La información obtenida con el sistema de control revela solamente el desempeño de lo que se mide (aunque no pueda compararse con el desempeño pasado y algún nivel planeado de [objetivos de] desempeño presentes o proyectados). Los datos del control no revelan por qué ocurrió ese nivel de desempeño, ni ofrecen explicaciones causales sobre los cambios en el desempeño de un período de informe a otro, o de un sitio a otro. Esta información proviene de un sistema de evaluación.

Un sistema de evaluación tiene una función complementaria pero distinta dentro de un marco de gestión basada en resultados. Un sistema de evaluación permite:

- Un estudio más profundo de los efectos directos y los impactos basados en resultados.

- Usar fuentes de datos distintas de los indicadores a los que se les hace seguimiento.
- Examinar factores demasiado difíciles o costosos de controlar en forma continua.
- Investigar por qué y cómo las tendencias a las que se les hace seguimiento con datos de control se mueven en la dirección que lo hacen.

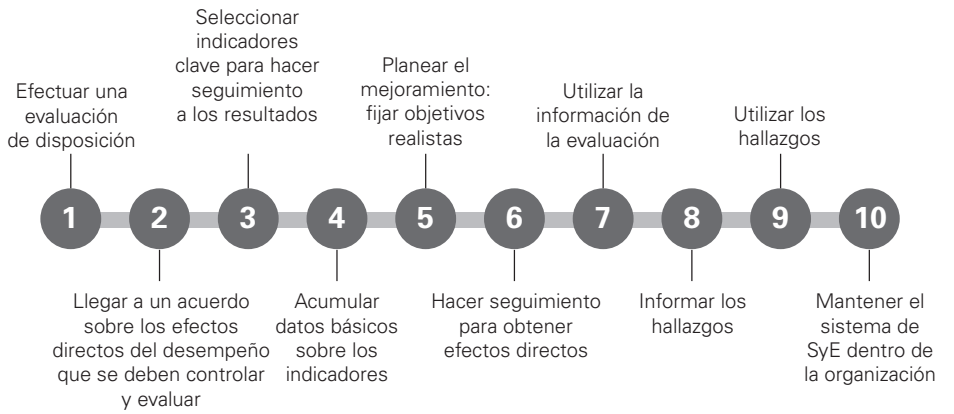
Los datos sobre impacto y la atribución causal no deben tomarse a la ligera, porque pueden tener un papel importante en la asignación estratégica de los recursos.

Diez pasos para construir un sistema de seguimiento y evaluación basado en resultados

La construcción de un sistema de SyE basado en resultados incluye 10 pasos (gráfico 3.3):

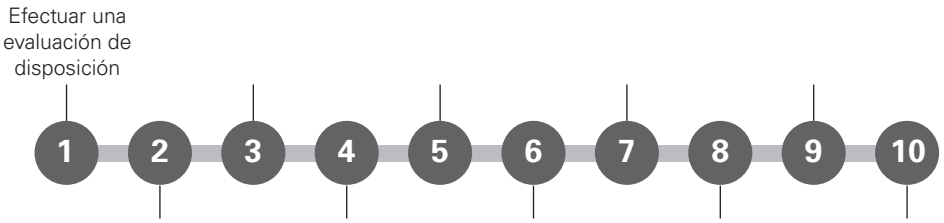
1. Efectuar una evaluación de disposición.
2. Llegar a un acuerdo sobre los efectos directos del desempeño que se deben controlar y evaluar.
3. Seleccionar los indicadores clave para hacer seguimiento a los resultados.
4. Acumular datos básicos sobre los indicadores.
5. Planear el mejoramiento: fijar objetivos realistas.
6. Hacer seguimiento para obtener resultados.
7. Utilizar la información de la evaluación.
8. Informar los hallazgos.
9. Utilizar los hallazgos.
10. Mantener el sistema de SyE dentro de la organización.

Gráfico 3.3 *Diez pasos para diseñar, construir y mantener un sistema de seguimiento y evaluación basado en resultados*



Fuente: Kusek y Rist, 2004.

Paso 1: Efectuar una evaluación de disposición



Una evaluación de disposición es una forma de determinar la capacidad y la disposición de un gobierno y sus asociados para el desarrollo, a fin de construir un sistema de SyE basado en resultados. En esta evaluación se abordan temas como la presencia o la falta de líderes, incentivos, papeles y responsabilidades, capacidad organizativa y barreras que impiden iniciar el sistema.

Incentivos

La primera parte de la evaluación de disposición incluye conocer qué incentivos existen para adelantar la construcción del sistema de SyE (y qué desincentivos pueden obstaculizar el progreso). Las preguntas que deben formularse son:

- ¿Qué impulsa la necesidad de construir un sistema de SyE?
- ¿Quiénes son los líderes para construir y usar un sistema de SyE?
- ¿Qué motiva a los líderes a construir el sistema de SyE?
- ¿Quién se beneficiará con el sistema?
- ¿Quién no se beneficiará?

Papeles y responsabilidades

Después es trascendental identificar quién es en realidad responsable de producir los datos en la organización y en otras organizaciones, y quiénes son los principales usuarios de los datos. Las preguntas que deben formularse son:

- ¿Cuáles son los papeles de los ministerios centrales y de línea en la evaluación del desempeño?
- ¿Cuál es el papel de la legislatura?
- ¿Cuál es el papel de la agencia principal de auditoría?
- ¿Comparten entre sí información los ministerios y agencias?
- ¿Existe una agenda política subyacente a los datos producidos?
- ¿Quién produce los datos en el país?
- ¿En qué niveles diferentes del gobierno se utilizan los datos?

Capacidad organizativa

Un elemento clave que impulsa la disposición de la organización para un sistema de SyE basado en resultados se relaciona con las destrezas, los recursos y experiencias disponibles para la organización. Las preguntas que deben formularse son:

- ¿Quién tiene las destrezas técnicas en la organización para diseñar e implementar un sistema de SyE?
- ¿Quién tiene las destrezas para administrar el sistema?
- ¿Qué sistemas de datos existen actualmente dentro de la organización y de qué calidad?
- ¿Qué tecnología hay disponible para apoyar el sistema de datos necesario? La capacidad de base de datos, disponibilidad de análisis de datos, *software* de presentación de información, etcétera, debería formar parte de la evaluación.
- ¿Qué recursos fiscales están disponibles para diseñar e implementar un sistema de SyE?
- ¿Qué experiencia tiene la organización con sistemas de presentación de información sobre desempeño?

Barreras

Al igual que con cualquier cambio organizativo, es crucial considerar qué podría interponerse en el camino de una implementación eficaz. Las preguntas que deben formularse son:

- ¿Se carece de recursos fiscales, voluntad política, un líder para el sistema, estrategia vinculada a los resultados, o experiencia?
- Si es así, ¿cómo pueden superarse tales barreras?

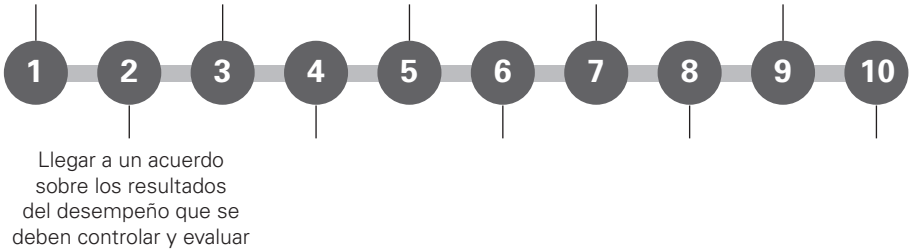
Las buenas prácticas sugieren que el éxito para establecer un sistema eficaz de SyE depende de varios factores, que incluye los siguientes:

- Un mandato claro para el SyE a nivel nacional.
- Documentos de estrategia de reducción de la pobreza, leyes y regulaciones.
- Fuerte liderazgo y apoyo en la mayoría de los niveles superiores del gobierno.
- Información confiable que pueda utilizarse para tomar decisiones en política y administración.
- Una sociedad civil que trabaje como asociada del gobierno en la formación y el seguimiento de la información sobre el desempeño.
- Focos de innovación que puedan servir de prácticas iniciales o programas piloto.

Al final de la evaluación de disposición, los funcionarios superiores del gobierno confrontan la cuestión de si continuar la construcción del sistema de SyE basado en

resultados. Esencialmente la pregunta es: "¿Continuar o no continuar?" (ya, pronto, o quizá más tarde).

Paso 2: Llegar a un acuerdo sobre los resultados del desempeño que se deben controlar y evaluar



Es importante generar interés al evaluar los resultados y los impactos que la organización o el gobierno intenta lograr, en lugar de simplemente concentrarse en las cuestiones de implementación (insumos, actividades y productos). Después de todo, los resultados (como mejorar la cobertura de los programas preescolares o mejorar el aprendizaje de los estudiantes de primaria) son lo que revela si se han obtenido, o no, los beneficios específicos.

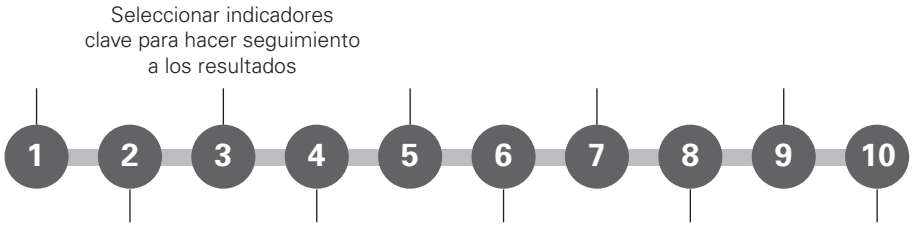
Los resultados y los impactos estratégicos deben concentrarse y dirigir la asignación de los recursos y las actividades del gobierno y sus asociados en el desarrollo. Estos impactos deben derivarse de las prioridades estratégicas del país. Los temas que deben considerarse al generar una lista de resultados incluyen los siguientes:

- ¿Existen metas nacionales o sectoriales declaradas (por ejemplo, Visión 2016)?
- ¿Se han hecho promesas políticas que especifiquen un mejor desempeño en un campo en particular?
- ¿Indican los datos de las encuestas a ciudadanos (como las tarjetas de puntuación) preocupaciones específicas?
- ¿Se encuentran los préstamos de los donantes ligados a metas específicas?
- ¿Existe una legislación autorizada?
- ¿Ha hecho el gobierno un compromiso serio para lograr los ODM?

Llegar a un acuerdo sobre los resultados es un proceso político que requiere aceptación y acuerdo de las principales partes interesadas. Para conocer sus preocupaciones se utilizan sesiones de intercambio de ideas, entrevistas, grupos de discusión y encuestas.

Los resultados hacen explícitas las consecuencias deseadas de una acción ("debes saber a dónde vas antes de empezar a andar"). Ellas representan los resultados que el cliente espera obtener. Antes de poder obtenerlos, deben transformarse en un conjunto de indicadores básicos. Fijar con claridad resultados –y derivar indicadores basados en ellos– es esencial para diseñar y construir sistemas de SyE basados en resultados.

Paso 3: Seleccionar indicadores clave para hacer seguimiento a los resultados



Lo que se mide se hace. La especificación de exactamente qué se va a medir para calibrar el logro de los resultados, no sólo ayuda a hacerle seguimiento al progreso, sino que también puede constituir una herramienta poderosa para enfocar los esfuerzos y crear sincronización dentro de una organización, si se hace con suficiente prontitud en el proceso.

Un indicador es una medida que, al hacerse un seguimiento sistemático a través del tiempo, indica progreso (o falta de progreso) en el cumplimiento de un objetivo. Responde a la pregunta "¿Cómo saber que hay éxito cuando se ve? En los nuevos sistemas de SyE todos los indicadores deberían ser cuantitativos, y los cualitativos pueden ir después, al madurar el sistema de SyE.

El desarrollo de indicadores es una actividad central en la construcción de un sistema de SyE, y dirige la posterior recopilación, el análisis y la presentación de datos. No deben subestimarse los problemas políticos y metodológicos para crear indicadores verosímiles y apropiados. Schiavo-Campo (1999) anota que los indicadores deberían ser:

- Claros (precisos e inequívocos).
- Pertinentes (apropiados al asunto al alcance de la mano).
- Económicos (disponibles a un costo razonable).
- Adecuados (capaces de suministrar base suficiente para evaluar un desempeño).
- Controlables (sensibles a la validación independiente).

A veces se pueden reducir costos al utilizar indicadores ya disponibles (aunque los evaluadores deberían ser conscientes de los riesgos de utilizarlos). Sin embargo, antes de hacerlo, es importante considerar cuán pertinentes son estos indicadores (y cuán pertinentes se perciben). Puede ser necesario adaptar o complementar algunos indicadores con otros más pertinentes al proyecto, programa o política que se evalúa.

El número de indicadores depende de la cantidad requerida para responder la pregunta "¿Se ha obtenido el resultado?" Y debería ser entre dos y siete. Una vez seleccionados, estos indicadores no se mantienen inalterables. Pueden agregarse nuevos indicadores y abandonarse algunos existentes a medida que se moderniza el sistema de control y mejora con el tiempo.

Los indicadores de desempeño seleccionados y las estrategias de recopilación de datos utilizadas para capturar información sobre estos indicadores deben estar basados en la realidad (Kusek y Rist, 2004). Los factores que deben considerarse son:

- Qué sistemas de datos están en funcionamiento.
- Qué datos pueden producirse actualmente.
- Qué capacidad existe para expandir las dimensiones de la recopilación y el análisis de los datos.

Al llenar cada celda de la matriz que se presenta en el cuadro 3.2, se obtiene una idea de la factibilidad de desplegar efectivamente cada indicador. En el cuadro 3.3 se presentan algunos ejemplos de indicadores.

Cuadro 3.2 *Matriz para la selección de indicadores*

Indicador	*Fuente de datos*	*Método de recopilación de datos*	*¿Quién recopilará los datos?*	*Frecuencia de la recopilación de datos*	*Costo de la recopilación de datos*	*Dificultad de la recopilación de datos*	*¿Quién analizará e informará sobre los datos?*	*¿Quién utilizará los datos?*
1.								
2.								
3.								

Fuente: Kusek y Rist, 2004.

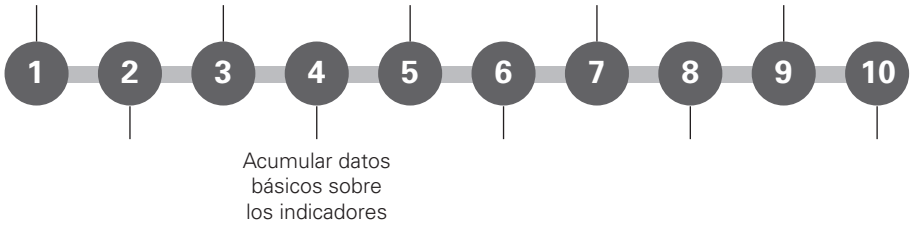
Cuadro 3.3 *Ejemplo del marco de desempeño: resultados e indicadores*

Resultado	*Indicadores*	*Datos básicos*	*Objetivos*
Mejor cobertura de programas preescolares	Porcentaje de niños urbanos elegibles inscritos en preescolar		
	Porcentaje de niños rurales elegibles inscritos en preescolar		
Mejores resultados de aprendizaje en primaria	Porcentaje de estudiantes de 6º grado con puntaje de 70% o mejor en las pruebas estandarizadas de matemáticas y ciencias		
	Porcentaje de estudiantes de 6º grado con puntaje más alto en las pruebas estandarizadas de matemáticas y ciencias comparado con los datos básicos		

Fuente: Kusek y Rist, 2004.

- Los evaluadores deben desarrollar sus propios indicadores para satisfacer las necesidades de la evaluación que realizan.
- El desarrollo de buenos indicadores requiere más de un intento y toma tiempo llegar a indicadores finales.
- Todos los indicadores deben expresarse en forma neutral, no como "incremento en" o "decremento en".
- Los evaluadores deben trabajar con proyectos piloto una y otra, y otra vez.

Paso 4: Acumular datos básicos sobre los indicadores



La medición del progreso (o la falta de progreso) en la obtención de resultados empieza con la descripción y la medición de las condiciones iniciales. Recopilar datos básicos significa esencialmente tomar las primeras medidas de los indicadores para averiguar "¿Dónde nos encontramos hoy?"

Un dato básico de desempeño suministra información (cualitativa o cuantitativa) sobre el desempeño en un indicador al principio (o inmediatamente antes) de la intervención. En realidad, una de las consideraciones al seleccionar los indicadores es la disponibilidad de datos básicos, que permite hacerle seguimiento al desempeño relativo de éstos.

Las fuentes de datos básicos pueden ser primarias (recopilados específicamente para este sistema de medición), o secundarias (recopilados con otro fin). Los datos secundarios pueden provenir del interior de una organización, del gobierno o de fuentes internacionales de datos. Se pueden obtener ahorros en dinero al utilizar estos datos, siempre que realmente suministren la información requerida. Es extremamente difícil devolverse y obtener datos primarios básicos si más tarde se encuentra que la fuente secundaria no satisface las necesidades de la evaluación.

Las posibles fuentes de datos básicos incluyen las siguientes:

- Registros escritos (en papel y también electrónicos).
- Participantes en un proyecto, programa o política.
- El público.
- Observadores entrenados.
- Mediciones y pruebas mecánicas.
- Sistemas de información geográfica.

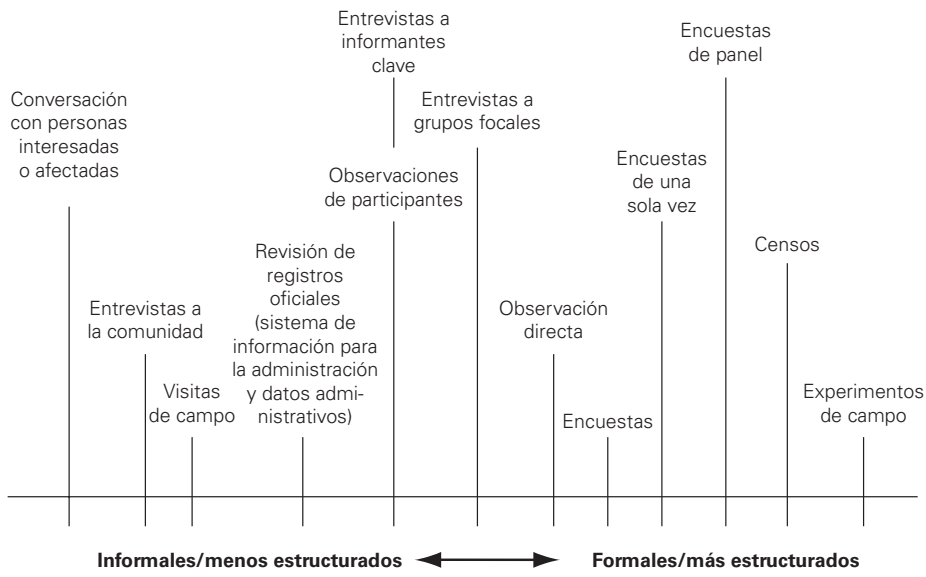
Una vez hayan seleccionado las fuentes de datos básicos para los indicadores, los evaluadores deciden quién debe recopilar los datos y cómo, e identifican y desarrollan instrumentos de recopilación de datos, como formularios para acumular información de archivos o registros, protocolos de entrevistas, encuestas e instrumentos de observación. Al desarrollar los instrumentos para la recopilación, tienen en cuenta cuestiones prácticas como las siguientes:

- ¿Existen buenos datos disponibles (o de fácil acceso) en la actualidad?
- ¿Pueden obtenerse los datos regular y oportunamente, de modo que pueda efectuarse un seguimiento del progreso?
- ¿Es factible y efectiva en costos la recopilación planeada de datos primarios?

Existen muchas formas de recopilar datos (según se verá en el capítulo 9), y pueden clasificarse desde las menos rigurosas, formales y costosas hasta las más rigurosas, formales y costosas (gráfico 3.4).

En el cuadro 3.4 se presenta el tercer paso en el desarrollo de resultados para la política educativa: establecer datos básicos.

Gráfico 3.4 *Espectro de los métodos de recopilación de datos*



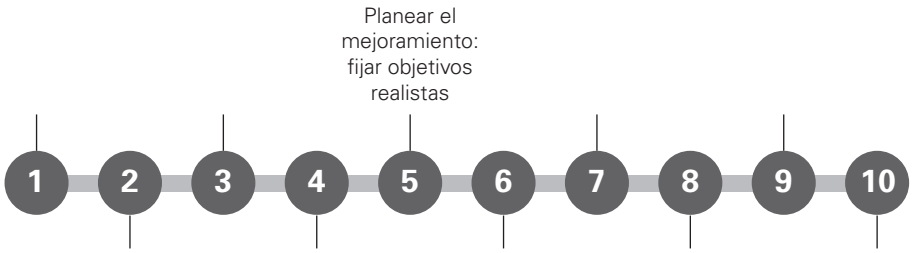
Fuente: Kusek y Rist, 2004.

Cuadro 3.4 *Ejemplo de marco de desempeño: resultados, indicadores y datos básicos*

Resultados	*Indicadores*	*Datos básicos*	*Objetivos*
Mejor cobertura de programas preescolares	Porcentaje de niños urbanos elegibles inscritos en preescolar	75% en zonas urbanas en 1999	
	Porcentaje de niños rurales elegibles inscritos en preescolar	40% en zonas rurales en 2000	
Mejores resultados de aprendizaje en primaria	Porcentaje de estudiantes de 6º grado con puntaje de 70% o mejor en las pruebas estandarizadas de matemáticas y ciencias	En 2002, 47% de los estudiantes obtuvieron un puntaje del 70% o más en matemáticas y del 50% o más en ciencias.	
	Porcentaje de estudiantes de 6º grado con puntaje más alto en las pruebas estandarizadas de matemáticas y ciencias comparado con los datos de la línea base	En 2002, el puntaje promedio para los estudiantes de 6º grado fue de 68% en matemáticas y de 53% en ciencias.	

Fuente: Kusek y Rist, 2004.

Paso 5: Planear el mejoramiento: fijar objetivos realistas



El siguiente paso –establecer objetivos– es el último en la formación del marco para el desempeño. Según Kusek y Rist (2004, p. 91), "En esencia, los objetivos son niveles cuantificables de los indicadores que un país, sociedad u organización desea alcanzar para una fecha dada".

La mayor parte de los resultados y casi todos los impactos en el desarrollo internacional son complejos y solamente ocurren en el largo plazo. Por consiguiente, es necesario establecer objetivos intermedios que especifiquen la cantidad de progreso que debe hacerse para el cumplimiento de un objetivo, en cuánto tiempo y con qué nivel de asignación de recursos. Medir el desempeño frente a estos objetivos puede implicar

el establecimiento de indicadores directos y de reemplazo, y también el uso de datos cuantitativos y cualitativos.

Pueden considerarse los impactos de la teoría del cambio como las metas a largo plazo que, en último término, se esfuerza por lograr la intervención. Los resultados constituyen un conjunto de objetivos secuenciales y factibles (relativos a los datos básicos) para los indicadores que se espera lograr durante todo el trayecto, en un tiempo especificado y realista (político y presupuestario). Expresado en otra forma, si una organización cumple sus objetivos a través del tiempo, obtendrá sus resultados (siempre que cuente con una buena teoría del cambio y la implemente en forma exitosa).

Al fijar objetivos para los indicadores es importante tener un claro conocimiento de lo siguiente:

- El punto de partida de los datos básicos (por ejemplo, el promedio de los tres últimos años, el último año, la tendencia promedio).
- Una teoría del cambio y una forma de desagregarla en un conjunto de logros limitados por el tiempo.
- Los niveles de recursos de financiación y de personal en el tiempo previsto para cumplir el objetivo.
- La cantidad de recursos externos que se espera complementen los recursos corrientes del programa.
- Las preocupaciones de las políticas pertinentes.
- La experiencia organizativa (especialmente administrativa) en la ejecución de proyectos y programas en este campo sustantivo.

En el gráfico 3.5 se muestra cómo identificar los objetivos que deben cumplirse como un eslabón de una cadena que, con el tiempo, lleva a obtener un resultado.

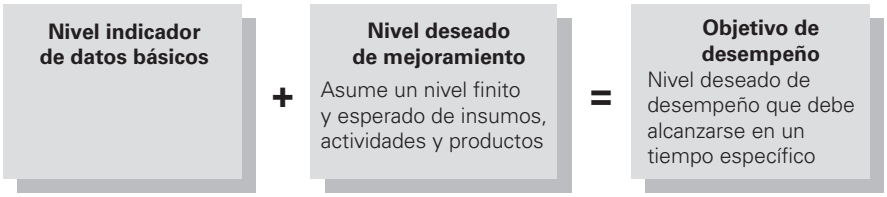
Sólo debe fijarse un objetivo para cada indicador. Si no se ha utilizado anteriormente un indicador, el evaluador debe tener precaución al fijar un objetivo específico y establecer más bien un rango. Los objetivos deben fijarse para el mediano plazo (no más de tres años), e incluir tiempos realistas para su logro. La mayor parte de los objetivos se fija anualmente, pero algunos pueden fijarse trimestralmente, o para períodos más largos.

El cuadro 3.5 muestra el último paso en el desarrollo de resultados para la política educativa: fijar los objetivos para el desempeño. Esta matriz completa viene a ser ahora el marco para el desempeño. Define los resultados y proporciona un plan para determinar si el programa tuvo éxito en el logro de estos resultados. El marco define el diseño de un sistema de SyE basado en resultados que empezará a suministrar información sobre si se están cumpliendo los objetivos intermedios en el camino para obtener los resultados a largo plazo.

El marco ayuda a los evaluadores a diseñar la evaluación, y puede ayudar a los administradores con el presupuesto, la asignación de recursos, la contratación de personal y otras funciones. Los administradores deben consultar el marco con frecuencia para garantizar que el proyecto, el programa o la política se encamina hacia los resultados deseados.

001El camino.indd 111 08/11/2010 03:27:44 p.m.

Gráfico 3.5 *Identificar el nivel esperado o deseado de mejoramiento requiere la selección de objetivos de desempeño*



Fuente: Kusek y Rist, 2004.

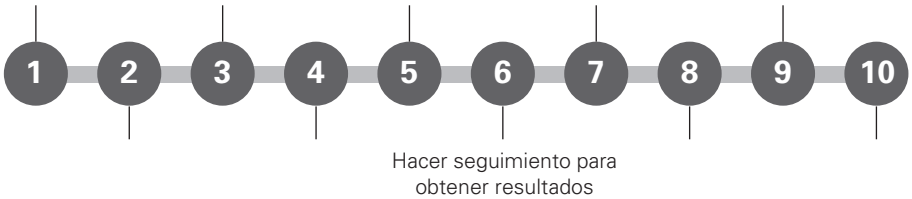
La focalización del desempeño es crítica para lograr los resultados de las políticas. Para el SyE basados en resultados es clave utilizar un proceso participativo y colaborador, que incluya indicadores básicos y niveles deseados de mejoramiento a través del tiempo.

Cuadro 3.5 *Ejemplo de marco de desempeño: resultados, indicadores, datos básicos y objetivos*

Resultados	*Indicadores*	*Datos básicos*	*Objetivos*
Mejor cobertura de programas preescolares	Porcentaje de niños urbanos elegibles inscritos en preescolar	75% en zonas urbanas en 1999	85% en zonas urbanas para 2006
	Porcentaje de niños rurales elegibles inscritos en preescolar	40% en zonas rurales en 2000	60% en zonas rurales para 2006
Mejores resultados de aprendizaje en primaria	Porcentaje de estudiantes de 6º grado con puntaje de 70% o mejor en las pruebas estandarizadas de matemáticas y ciencias.	En 2002, 47% de los estudiantes obtuvieron un puntaje del 70% o más en matemáticas y del 50% o más en ciencias.	Para 2006, el 80% de los estudiantes obtendrá un puntaje del 70% o más en matemáticas y el 67% uno de 70% o más en ciencias.
	Porcentaje de estudiantes de 6º grado con puntaje más alto en las pruebas estandarizadas de matemáticas y ciencias comparado con los datos de la línea base.	En 2002, el puntaje promedio para los estudiantes de 6º grado fue de 68% en matemáticas y de 53% en ciencias.	En 2006, el puntaje promedio de la prueba será de 78% en matemáticas y de 65% en ciencias.

Fuente: Kusek y Rist, 2004.

Paso 6: Hacer seguimiento para obtener resultados

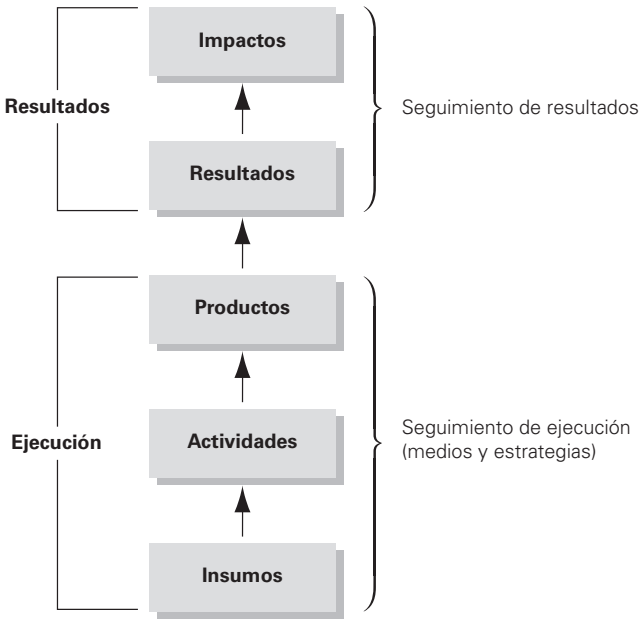


Un sistema de control basado en resultados hace seguimiento tanto a la ejecución (insumos, actividades, productos) como a los resultados (resultados e impactos). En el gráfico 3.6 se muestra cómo cada uno de estos tipos de seguimiento se ajusta al modelo.

Todo resultado tiene varios indicadores, cada uno de los cuales tiene un objetivo fijado. Debe haber una coordinación y un manejo de una serie de actividades y estrategias para obtener estos resultados (gráfico 3.7).

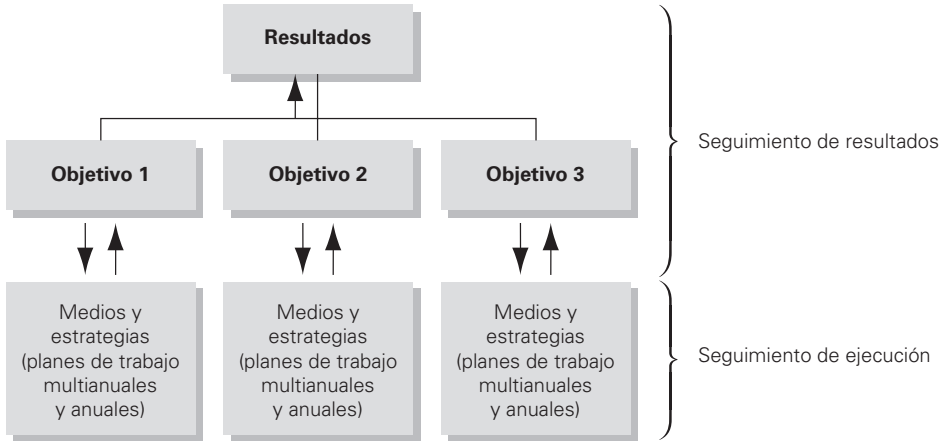
Es de importancia crucial vincular el seguimiento de la ejecución con el de los resultados. El gráfico 3.8 ofrece un ejemplo para reducir la mortalidad infantil.

Gráfico 3.6 *Tipos clave de seguimiento*



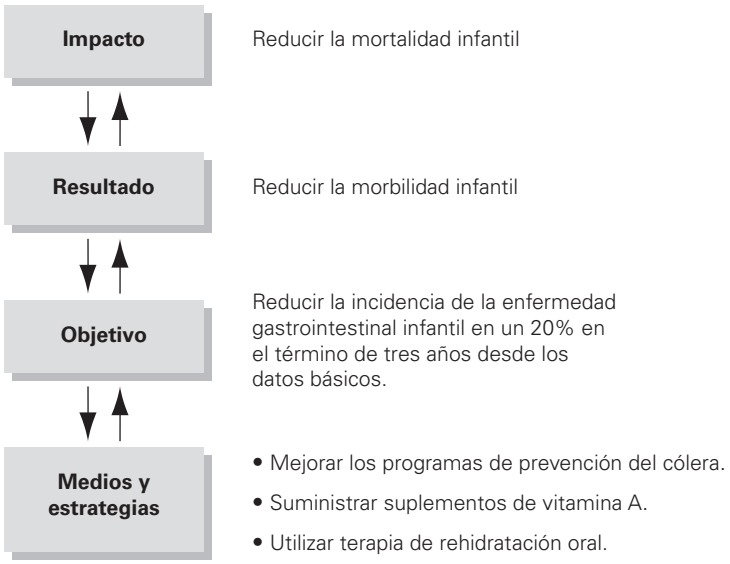
Fuente: Adaptado de Binnendijk, 2000.

Gráfico 3.7 *Vínculos entre el seguimiento de la ejecución y el de los resultados*



Fuente: Kusek y Rist, 2004.

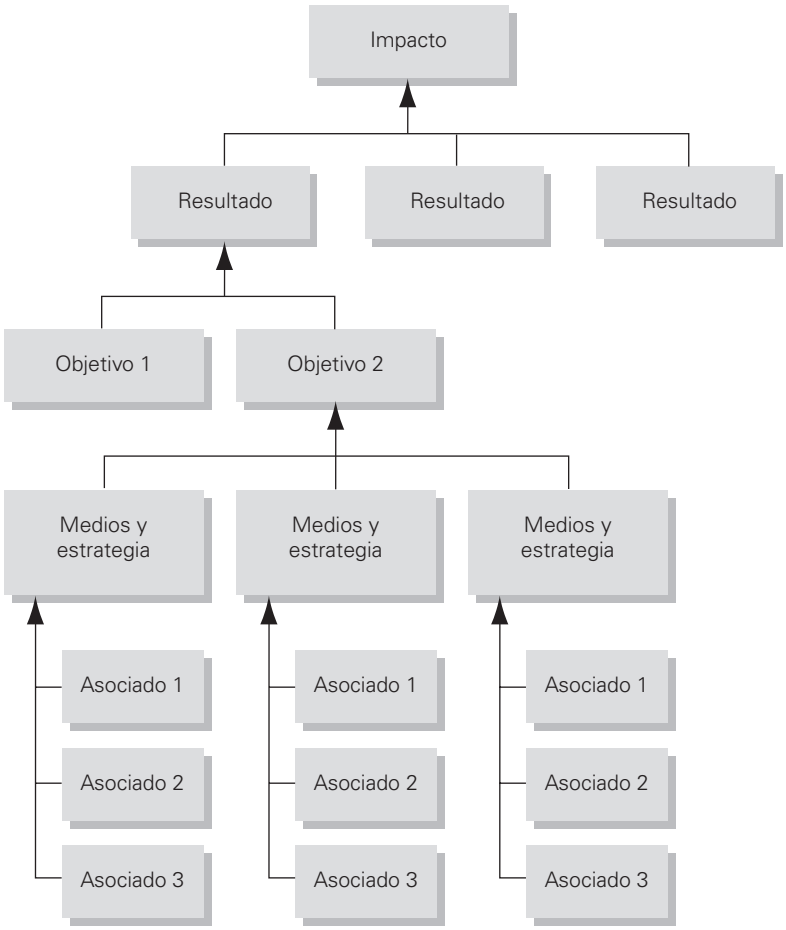
Gráfico 3.8 *Ejemplo de vínculo entre el seguimiento de ejecución y el de los resultados*



Fuente: Kusek y Rist, 2004.

El trabajo con asociados es cada vez más la norma en la labor para el desarrollo. Muchos asociados al nivel inferior de esta jerarquía contribuyen potencialmente con insumos, actividades y productos como parte de la estrategia para el logro de los objetivos (gráfico 3.9).

Gráfico 3.9 *El logro de resultados por medio de la asociación*



Fuente: Kusek y Rist, 2004.

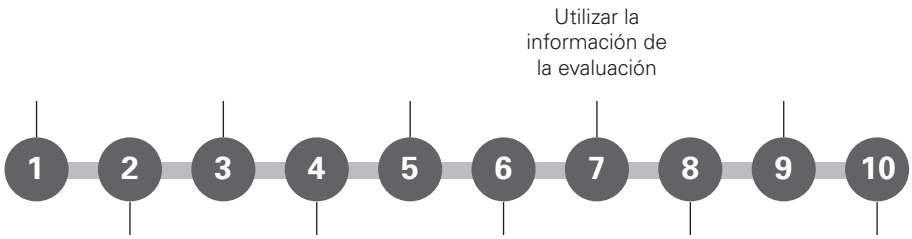
Un sistema fuerte de SyE debe tener el apoyo de un presupuesto, planes de contratación de personal y planeación de actividades. La construcción de un sistema eficaz de SyE incluye tareas administrativas e institucionales, como las siguientes:

- Establecer la recopilación de datos, su análisis y lineamientos para la información.
- Designar los responsables de las distintas actividades.
- Establecer medios de control de calidad.
- Fijar cronogramas y costos.
- Trabajar por medio de los papeles y responsabilidades del gobierno, otros asociados en el desarrollo y la sociedad civil.
- Fijar lineamientos sobre la transparencia y la divulgación de la información, y el análisis.

Para tener éxito, todo sistema de SyE necesita lo siguiente:

- Propiedad.
- Administración.
- Mantenimiento.
- Credibilidad.

Paso 7: Utilizar la información de la evaluación



La evaluación tiene un papel importante para complementar la información adquirida al controlar el progreso en obtener los resultados y los impactos. Mientras el control revela lo que se está haciendo en relación con los indicadores, objetivos y resultados, la evaluación revela si:

- Se están haciendo las cosas correctas (estrategia).
- Se están haciendo correctamente (operaciones).
- Existen mejores formas de hacerlas (aprendizaje).

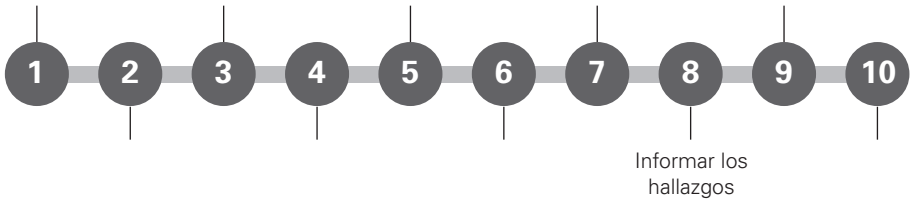
La evaluación atiende asuntos muy importantes que van más allá de un simple sistema de control. Por ejemplo, el diseño de muchas intervenciones se basa en ciertos supuestos causales sobre el problema o el asunto que se atiende. La evaluación puede confirmar u objetar estos supuestos causales, utilizando modelos de evaluación basados en la teoría y modelos de lógica (como se verá en el capítulo 4). La evaluación puede también explorar más a fondo un resultado o una tendencia interesante o problemática que surja del sistema de control (averiguar, por ejemplo, por qué las muchachas abandonan más temprano el colegio en una aldea que los muchachos).

Una evaluación debe utilizarse adicionalmente al seguimiento cuando:

- Exista un resultado o un desempeño inesperado y atípico que requiera investigación adicional.
- Se efectúen asignaciones de recursos o presupuesto entre proyectos, programas o políticas.
- Se discuta si expandir o no un programa piloto.
- Ha pasado un largo período sin mejora, sin que exista una explicación clara del porqué.
- Programas o políticas similares informan resultados divergentes, o cuando los indicadores para el mismo resultado presentan tendencias divergentes.
- Se intenta comprender el mérito, el valor y el significado de lo realizado.
- Se examinan detenidamente los costos en relación con los beneficios.

Para que los gobiernos y organizaciones puedan depender de la información recopilada de un sistema de SyE, deben contar con la calidad y la confiabilidad de la información que se recopila. La información deficiente, inexacta y sesgada no es útil para nadie.

Paso 8: Informar los hallazgos



Una parte crítica de SyE es determinar qué hallazgos se informan a quién, en qué formato y a qué intervalos. El análisis y la presentación de los datos:

- Suministran información sobre el estado de los proyectos, programas y políticas.
- Arrojan pistas sobre problemas.
- Crean oportunidades para considerar mejoras en las estrategias de ejecución.
- Suministran información importante a través del tiempo sobre tendencias y direcciones.
- Ayudan a confirmar u objetar la teoría del cambio subyacente al proyecto, programa o política. (El análisis y la presentación de datos se cubren en detalle en capítulos posteriores.)

El evaluador debe estar seguro de averiguar cuáles son los principales puntos de decisión a nivel de proyecto, programa o política, de manera que sea claro cuándo serán más útiles los hallazgos del SyE para quienes toman las decisiones. Si los datos y el análisis llegan demasiado tarde, no podrán afectar las decisiones.

Deben informarse todos los resultados importantes, sean positivos o negativos (cuadro 3.6). Un buen sistema de SyE debe proveer un sistema de advertencias tempranas para detectar problemas o inconsistencias y demostrar el valor de una intervención. Los informes sobre el desempeño deben incluir explicaciones sobre resultados deficientes o decepcionantes y documentar los pasos que se han dado para evitarlos.

Al analizar y presentar los datos, los evaluadores deberían:

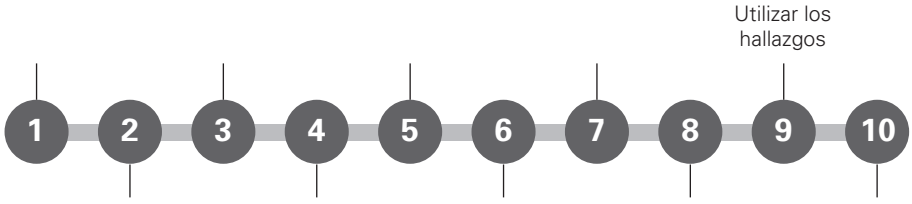
- Comparar los datos de los indicadores con los datos básicos y los objetivos y suministrar esta información en un gráfico de fácil comprensión (*ver* capítulo 13).
- Comparar la información actual con datos pasados y buscar patrones y tendencias.
- Ser cuidadosos al sacar conclusiones apresuradamente basadas en información insuficiente (cuantos más datos recopilados, más seguro puede estar el evaluador de que las tendencias son reales).
- Proteger al mensajero: no debe castigarse a las personas que portan malas noticias (los hallazgos incómodos pueden indicar nuevas tendencias o notificar prontamente a los administradores sobre problemas, dándoles tiempo para resolverlos).

Cuadro 3.6 *Cuadro de ejemplo de presentación de resultados*

Indicador de resultado	*Línea base*	*Actual*	*Objetivo*	*Diferencia (objetivo – actual)*
Tasa de hepatitis (*N* = 6.000)	30	35	20	–5
Porcentaje de niños con mejor estado general de salud (*N* = 9.000)	20	20	24	–4
Porcentaje de niños que presentan 4 de 5 posibles anotaciones positivas en los exámenes físicos (*N* = 3.500)	50	65	65	0
Porcentaje de niños con mejor estado de nutrición (*N* = 14.000)	80	85	83	2

Fuente: Kusek y Rist, 2004.

Paso 9: Utilizar los hallazgos



El punto crucial de un sistema de SyE no es simplemente generar información basada en los resultados, sino suministrar esa información a los usuarios apropiados en forma oportuna, de tal modo que la puedan tener en cuenta (o escoger ignorarla) para administrar el proyecto, el programa o la política. Los asociados en el desarrollo y la sociedad civil tienen papeles importantes en el uso de la información para fortalecer los procesos de rendición de cuentas, transparencia y asignación de recursos.

Las estrategias para compartir información, que pueden implementarse a cualquier nivel de gobierno, incluyen las siguientes:

- *Apoderar a los medios.* Los medios contribuyen a difundir los hallazgos generados por los sistemas de SyE basados en resultados, y son también útiles para exponer la corrupción y exigir una mejor gobernabilidad.
- *Aprobar legislación de libertad de información.* La libertad de información es una poderosa herramienta que puede emplearse para compartir información con las partes interesadas preocupadas.
- *Instituir el e-government.* El *e-government* implica usar tecnologías de información para proveer una mejor accesibilidad, información de alcance y servicios. El *e-government* permite a las partes interesadas interactuar directamente con el gobierno para obtener información y aun realizar negocios en línea.
- *Agregar información sobre archivos internos y externos en Internet.* La información puede compartirse publicándola en sitios web internos (de la agencia o del gobierno) y externos, y hacer lo mismo con los hallazgos de desempeño publicados. Muchas agencias están desarrollando también bases de datos consultables para los hallazgos de SyE.
- *Publicar informes anuales de presupuesto.* La mejor forma de comunicar la forma como se emplea el dinero del contribuyente es publicar el presupuesto, pues esto permite que los ciudadanos observen la calidad y el nivel de los servicios que presta el gobierno y la prioridad que éste otorga a servicios o a programas particulares.
- *Vincular a la sociedad civil con grupos de ciudadanos.* Esto estimula a los grupos a orientarse más hacia la acción, a ser más responsables y a tener mejor disposición a acordar sobre la información que necesitan.
- *Fortalecer la fiscalización legislativa.* Las legislaturas de muchos países, tanto desarrollados como en desarrollo, solicitan información sobre el desempeño como parte de sus funciones de fiscalización y buscan verificar que los presupuestos se utilicen con eficacia.
- *Fortalecer el despacho del auditor general.* El despacho del auditor general es un asociado clave para determinar cuán eficazmente está funcionando el gobierno. Cuando las agencias de auditoría demandan más información sobre el desempeño del sector público, los proyectos, programas y políticas se implementan en forma más eficaz.
- *Compartir y comparar hallazgos con los asociados para el desarrollo.* Como resultado de introducir estrategias nacionales de reducción de la pobreza y

estrategias y políticas similares, los asociados para el desarrollo (especialmente las agencias de ayuda bilateral y multilateral) comparten y comparan los resultados y hallazgos.

Comprender la utilidad de la información sobre el desempeño para varios usuarios es una razón básica para construir un sistema de SyE. Los usuarios potenciales clave en muchas sociedades, como son los ciudadanos, las ONG y el sector privado, se dejan por fuera del flujo de información. Los datos del SyE tienen usos tanto internos (gubernamentales) como externos (de la sociedad) que hay que reconocer y legitimar (recuadro 3.3).

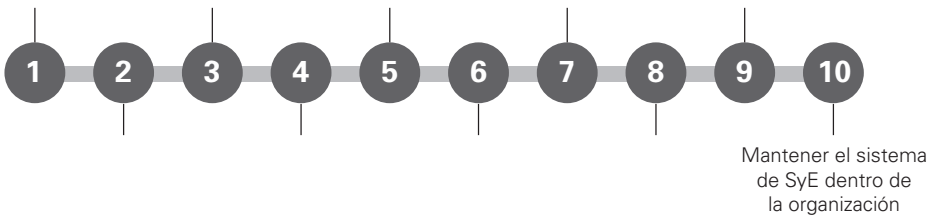
Recuadro 3.3 *Diez usos de los resultados obtenidos*

Los resultados obtenidos pueden utilizarse para:

1. Responder a las demandas de rendición de cuentas de los funcionarios electos y el público.
2. Contribuir a formular y justificar las solicitudes de presupuesto.
3. Tomar decisiones operativas sobre asignación de recursos.
4. Activar exámenes profundos sobre los problemas que existen en el desempeño (con la teoría del cambio o la implementación) y las correcciones necesarias.
5. Motivar al personal para continuar mejorando los programas.
6. Controlar el desempeño de los contratistas y concesionarios (ya no es suficiente que documenten lo ocupados que se encuentran).
7. Proporcionar datos para evaluaciones especiales y profundas de los programas.
8. Controlar la prestación de los servicios frente a objetivos precisos de resultados (¿se están haciendo las cosas bien?)
9. Apoyar esfuerzos estratégicos y otros esfuerzos de planeación a largo plazo (¿se están haciendo las cosas que deben hacerse?)
10. Comunicarse con el público para generar confianza pública.

Fuente: Hatry, 1999.

Paso 10: Mantener el sistema de SyE dentro de la organización



Asegurar la longevidad y la utilidad de un sistema de SyE basado en resultados constituye un reto. Para la sostenibilidad son de importancia crucial estos seis componentes:

- Demanda.
- Claridad y responsabilidad en los papeles.
- Información digna de confianza y verosímil.
- Rendición de cuentas.
- Capacidad.
- Incentivos.

Cada componente debe atenderse continuamente para asegurar la viabilidad del sistema.

Demanda

Existen varias formas de generar y mantener la demanda de SyE:

- Construir una estructura formal que requiera presentación de información regularmente sobre los resultados del desempeño (por ejemplo, un requerimiento de presentación anual de informe para las unidades organizativas).
- Difundir la disponibilidad de esta información a través de los medios, generando con esto una demanda de los organismos oficiales, grupos de ciudadanos, donantes y el público.
- Convertir en una práctica la traducción de la estrategia a metas y objetivos específicos, de modo que quienes están interesados en la dirección estratégica de la organización puedan seguir el progreso del cumplimiento de esas metas.

Claridad y responsabilidad en los papeles

Una de las intervenciones estructurales más importantes para institucionalizar un sistema de SyE es la creación de líneas claras y formales de autoridad, y responsabilidades para recopilar, analizar y presentar información sobre el desempeño. Para hacer esto es necesario:

- Emitir una orientación clara sobre quién es responsable de cuáles componentes del sistema de SyE, y hacer que los individuos sean responsables en sus revisiones sobre el desempeño.
- Construir un sistema que vincule las funciones de planificación central y finanzas con las de línea o de sector para estimular un enlace entre los ciclos de asignación presupuestaria y el suministro de información de SyE, esencialmente un sistema presupuestario basado en el desempeño.
- Construir un sistema en el que exista una demanda de información en cada uno de los niveles del sistema (es decir, que no haya parte del sistema en que la información simplemente pase sin que se utilice).

001El camino.indd 121 08/11/2010 03:27:45 p.m.

Información digna de confianza y verosímil

El sistema de información sobre el desempeño debe poder producir tanto buenas como malas noticias. En consecuencia, los productores de la información deben estar protegidos de represalias políticas. La información que el sistema produce debe ser transparente y estar sujeta a verificación independiente (por ejemplo, una revisión de la auditoría general de la nación del gobierno, o de un grupo independiente de profesores universitarios).

Rendición de cuentas

Se deben encontrar formas de compartir información transparente con las partes interesadas externas que tienen interés en el desempeño. Los grupos de interesados clave para esto incluyen organizaciones de la sociedad civil, los medios de comunicación, el sector privado y el gobierno.

Capacidad

Una de las primeras cosas que deben considerarse al construir un sistema de SyE es la de iniciar una evaluación de disposición y un enfoque en la capacidad organizativa. Los elementos clave con los que debe contarse incluyen las destrezas técnicas sólidas en recopilación y análisis de datos, las destrezas empresariales en el establecimiento de metas estratégicas y el desarrollo organizativo, los sistemas existentes de recopilación y recuperación de datos, la disponibilidad corriente de recursos financieros y la experiencia institucional en seguimiento y evaluación.

Incentivos

Se deben introducir incentivos para alentar el uso de la información sobre el desempeño. Se debe reconocer y recompensar el éxito, y abordar el tratamiento de los problemas. No debe castigarse a los mensajeros, hay que valorar el aprendizaje organizativo y compartir los ahorros en el presupuesto. No se debe depender de sistemas corruptos o ineficaces para producir información y análisis de calidad.

Conclusiones

No se requiere que la construcción de un sistema de SyE deba hacerse de acuerdo con estos 10 pasos, y pueden desarrollarse estrategias con mayor o menor número de pasos. El reto es garantizar que las funciones y actividades se reconozcan, se agrupen en formas lógicas y se implementen en la secuencia apropiada.

Los sistemas de SyE basados en resultados constituyen poderosas herramientas administrativas. Contribuyen a formar y fomentar el cambio en la forma como operan los gobiernos y las organizaciones, y también a construir una base de conocimiento sobre lo que funciona bien y lo que no.

001El camino.indd 122 08/11/2010 03:27:45 p.m.

Un sistema de SyE basado en resultados debe ser iterativo y recibir atención, recursos y compromiso político continuos. El desarrollo del cambio cultural hacia la orientación a resultados es algo que toma tiempo, pero el tiempo, el esfuerzo y las recompensas bien valen la pena hacer el intento.

La demanda de formación de capacidad nunca termina; la única forma como una organización puede deslizarse es cuesta abajo, y varios pasos ayudan a asegurar la configuración y el sostenimiento de un sistema de SyE:

- Mantenga a su lado los líderes y ayúdeles.
- Convenza al Ministerio de Finanzas y a la legislatura de que un sistema de SyE necesita recursos sostenidos, en igual forma que un sistema de presupuesto (el volumen de recursos asignados a un sistema de SyE debe ser aproximadamente igual al asignado al sistema presupuestario).
- Busque cualquier oportunidad para vincular la información de los resultados al presupuesto y a las decisiones sobre asignación de recursos.
- Empiece con programas piloto para demostrar un SyE eficaces basados en resultados.
- Empiece con una estrategia de enclave (construida, por ejemplo, alrededor de islas de innovación), y no con el enfoque que abarca la totalidad del gobierno.
- Controle el progreso hacia la ejecución y la obtención de resultados.
- Complemente el seguimiento del desempeño con evaluaciones para asegurar un mejor conocimiento de los resultados del sector público.

Una vez se ha desarrollado el marco para una evaluación (paso 7), puede utilizarse para construir una teoría del cambio, elegir un enfoque, empezar a escribir las preguntas y escoger un diseño para la evaluación. Estos temas se cubren en los capítulos 4 a 7.

Resumen

Un sistema de SyE basado en resultados puede ser una herramienta valiosa para asistir a quienes formulan las políticas y a quienes toman las decisiones a hacer seguimiento de los resultados y los impactos de los proyectos, programas y políticas. A diferencia de la evaluación tradicional, el SyE basado en resultados van más allá de un énfasis en los insumos y productos para concentrarse en los resultados y los impactos. Es la arquitectura clave para cualquier sistema de medición del desempeño.

Los sistemas de SyE basados en resultados:

- Utilizan datos básicos para describir un problema antes de empezar una iniciativa.
- Hacen seguimiento a los indicadores de los resultados que deben obtenerse.
- Recopilan datos sobre insumos, actividades y productos, y su contribución a la obtención de los resultados.
- Evalúan la fortaleza y la propiedad de la teoría del cambio utilizada.

- Incluyen información sistemática para las partes interesadas.
- Se desarrollan con asociados estratégicos.
- Capturan información sobre el éxito o el fracaso de la estrategia de asociación en el logro de los resultados deseados.
- Se esfuerzan constantemente para suministrar información verosímil y útil como herramienta de gestión.

Para diseñar y desarrollar un sistema de SyE basado en resultados se recomiendan 10 pasos:

1. Efectuar una evaluación de disposición.
2. Llegar a un acuerdo sobre los resultados del desempeño que se deben controlar y evaluar.
3. Seleccionar indicadores clave para hacer seguimiento a los resultados.
4. Acumular datos básicos sobre los indicadores.
5. Planear el mejoramiento: fijar objetivos realistas.
6. Hacer seguimiento para obtener resultados.
7. Utilizar la información de la evaluación.
8. Informar los hallazgos.
9. Utilizar los hallazgos.
10. Mantener el sistema de SyE dentro de la organización.

No es fácil desarrollar y mantener un sistema de SyE basado en resultados, y se requieren compromiso continuo, líderes, tiempo, esfuerzo y recursos. Puede haber problemas organizativos, técnicos y políticos. Pueden necesitarse varias revisiones al sistema original para adecuarlo de forma que satisfaga las necesidades de la organización, pero vale la pena el esfuerzo de hacerlas.

Actividades del capítulo 3

Ejercicio de aplicación 3.1: trabaje con la lógica correcta

¿Qué tan dispuesta está su organización para diseñar e implementar un sistema de SyE basado en resultados? Califique su organización en cada una de las siguientes dimensiones y explique con comentarios su calificación. Discuta con un colega las barreras a la implementación y cómo pueden resolverse.

1. Incentivos (encierre en un círculo la calificación apropiada):
   muchos incentivos pocos incentivos varios desincentivos
   Comentarios:
   Estrategias para mejoramiento:

2. Papeles y responsabilidades (encierre en un círculo la calificación apropiada):

   muy claros algo claros poco claros

   Comentarios:

   Estrategias para mejoramiento:

3. Capacidad organizativa (encierre en un círculo la calificación apropiada):

   excelente adecuada débil

   Comentarios:

   Estrategias para mejoramiento:

4. Barreras (encierre en un círculo la calificación apropiada):

   sin barreras serias muy pocas barreras barreras serias

   Comentarios:

   Estrategias para mejoramiento:

Ejercicio de aplicación 3.2: identificar insumos, actividades, productos, resultados e impactos

Identifique si cada una de las afirmaciones siguientes es un insumo, una actividad, un producto, un resultado o un impacto a largo plazo. De ser posible, discuta con un colega y explique la base sobre la cual clasificó cada afirmación.

1. Las microempresas de propiedad de mujeres contribuyen de manera significativa a reducir la pobreza en las comunidades en las que operan.
2. El gobierno ha ofrecido fondos para préstamos a microempresas.
3. El gobierno aprobó 61 solicitudes de graduados de programas.
4. El Ministerio de Educación identificó instructores para cursos.
5. Setenta y dos mujeres terminaron la capacitación.
6. El ingreso de los graduados aumenta un 25% en el primer año después de terminar el curso.
7. Cien mujeres asistieron a capacitación en administración de microempresas.
8. Se suministra información sobre la disponibilidad de préstamos para programas de microempresas a las comunidades.

Ejercicio de aplicación 3.3: desarrollar indicadores

1. Identifique un programa o política con el que esté familiarizado. ¿Cuál es el impacto principal que intenta causar? ¿Cuáles son los dos resultados que usted esperaría ver si la intervención está en camino de causar ese impacto?

   Impacto: _____

   Resultado 1: _____

   Resultado 2: _____

2. Empezando con los resultados, identifique dos o tres indicadores que usted utilizaría para hacer seguimiento al progreso frente a cada uno.

   Resultado 1: _____

   - Indicador a: _____
   - Indicador b: _____
   - Indicador c: _____

   Resultado 2: _____

   - Indicador a: _____
   - Indicador b: _____
   - Indicador c: _____

   Impacto: _____

   - Indicador a: _____
   - Indicador b: _____
   - Indicador c: _____

Referencias y lecturas adicionales

Banco Mundial. 1997. *Informe sobre el Desarrollo Mundial: El Estado en un mundo cambiante.* Washington, DC.

Binnendijk, Annette. 2000. "Results-Based Management in the Development Cooperation Agencies: A Review of Experience". Documento preparado para el OECD/CAD Working Party on Aid Evaluation, París, febrero 10-11 (revisado en octubre, 2000).

Boyle, R. y D. Lemaire, eds. 1999. *Building Effective Evaluation Capacity.* Nueva Brunswick, NJ: Transaction Books.

Furubo, Jan-Eric, Ray C. Rist y Rolf Sandahl, eds. 2002. *International Atlas of Evaluation.* Nueva Brunswick, NJ: Transaction Books.

Hatry, Harry P. 1999. *Performance Measurement: Getting Results*. Washington, DC: Urban Institute Press.

IFAD (International Fund for Agriculture Development). 2002. *A Guide for Project M&E: Managing for Impact in Rural Development*. Roma. www.ifad.org/evaluation/guide/.

Khan, M. Adil. 2001. *A Guidebook on Results-Based Monitoring and Evaluation: Key Concepts, Issues and Applications*. Government of Sri Lanka, Ministry of Plan Implementation, Monitoring and Progress Review Division, Colombo.

Kusek, Jody Zall y Ray C. Rist. 2001. "Building a Performance-Based Monitoring and Evaluation System: The Challenges Facing Developing Countries". *Evaluation Journal of Australasia* 1 (2): 14-23.

———. 2003. "Readiness Assessment: Toward Performance Monitoring and Evaluation in the Kyrgyz Republic". *Japanese Journal of Evaluation Studies* 31 (1): 17-31.

———. 2004. *Ten Steps to Building a Results-Based Monitoring and Evaluation System.* Washington, DC: Banco Mundial.

Malik, Khalid y Christine Roth, eds. 1999. *Evaluation Capacity Development in Asia.* United Nations Development Programme Evaluation Office, Nueva York.

Osborn, David y Ted Gaebler. 1992. *Reinventing Government*. Boston: Addison-Wesley Publishing.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2002. *Glossary of Key Terms in Evaluation and Results-Based Management*. Development Co-operation Directorate and Development Assistance Committee. París.

Schiavo-Campo, Salvatore. 1999. "'Performance' in the Public Sector". *Asian Journal of Political Science* 7 (2): 75-87.

UNPF (Fondo de Población de las Naciones Unidas). 2002. *Monitoring and Evaluation Toolkit for Program Managers.* Office of Oversight and Evaluation. www.unfpa.org/monitoring/toolkit.htm.

Valadez, Joseph y Michael Bamberger. 1994. *Monitoring and Evaluation Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers.* Washington, DC: Banco Mundial.

Weiss, Carol. 1972. *Evaluation Research: Methods for Assessing Program Effectiveness.* Englewood Cliffs, NJ: Prentice Hall.

Wholey, Joseph S., Harry Hatry y Kathryn Newcomer. 2001. "Managing for Results: Roles for Evaluators in a New Management Era". *American Journal of Evaluation* 22 (3): 343-47.

Sitios web

Banco Mundial. *Core Welfare Indicators Questionnaire.* Washington, DC. http://www4.worldbank.org/afr/stats/cwiq.cfm.

———. 2001. *Tools: Capturing Experience Monitoring and Evaluation.* Upgrading Urban Communities Resource Framework, Washington, DC. http://web.mit.edu/urbanupgrading/upgrading/issues-tools/tools/monitoring-eval.html#Anchor-Monitoring-56567.

———. 2008. *Online Atlas of the Millennium Development Goals: Building a Better World*. Washington, DC. http://devdata.worldbank.org/atlas-mdg/.

CIID (Centro Internacional de Investigaciones para el Desarrollo). 2004. *Evaluation Planning in Program Initiatives.* Ottawa. http://web.idrc.ca/uploads/user-S/108549984812guideline-web.pdf.

IFAD (International Fund for Agricultural Development). *Practical Guide on Monitoring and Evaluation of Rural Development Projects.* http://www.ifad.org/ evaluation/oe/process/guide/index.htm.

Kellogg Foundation. 1998. *Evaluation Handbook.* http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf.

Sitio web especializado en control y evaluación. http://www.mande.co.uk/specialist.htm.

Uganda Communications Commission. 2005. "Monitoring and Evaluation". En *Funding and Implementing Universal Access: Innovation and Experience from Uganda.* CIID, Ottawa. http://www.idrc.ca/en/ev-88227-201-1-DO_TOPIC.html.

CAPÍTULO 4

Conocer el contexto de la evaluación y la teoría programática del cambio

Este es el primero de dos capítulos en los que se examina la planeación de la evaluación. En este capítulo se trata sobre el estado inicial de una evaluación: cómo empezar. Una evaluación que empieza con un diseño bien planeado tiene mayor probabilidad de terminar en el tiempo y con el presupuesto previsto y de satisfacer las necesidades del cliente y otras partes interesadas. En un análisis del estado inicial se investigan e identifican las enseñanzas del pasado, se confirma o se cuestiona la teoría subyacente al programa, y se establece el contexto que influye en el programa.

Análisis preliminar

¿Por dónde comenzar? Si se desea llegar al destino correcto, es mejor empezar averiguando qué dirección tomar y qué han aprendido ya otros sobre la ruta hacia ese destino. Será conveniente recopilar información crítica para tomar las decisiones sobre tiempo, costos, riesgos y procesos.

Un *análisis preliminar* es una investigación de un asunto o un problema para determinar qué se conoce acerca de él

Análisis preliminar: Investigación de un asunto o problema para determinar qué se conoce acerca de él, y cómo proceder al desarrollar un enfoque evaluativo para él.

y cómo proceder al desarrollar un enfoque evaluativo para él. Es lo que un evaluador hace para estimar qué debe hacer.

En un análisis preliminar el evaluador investiga los tipos de preguntas siguientes:

- ¿Quién es el principal cliente de la evaluación? ¿Quiénes son otras partes interesadas importantes? ¿Qué problemas identifican para la evaluación?
- ¿Cómo afectará a la evaluación la oportunidad de la evaluación en relación con la implementación del proyecto, programa o política?
- ¿Con cuánto tiempo se cuenta para completar la evaluación?
- ¿Qué naturaleza y cuantía tienen los recursos disponibles?
- ¿Tiene aplicabilidad la teoría de las ciencias sociales para la evaluación?
- ¿Qué han encontrado las evaluaciones de programas similares? ¿Qué problemas suscitaron?
- ¿Cuál es la teoría del cambio subyacente al proyecto, programa o política?
- ¿Qué datos existentes pueden utilizarse para esta evaluación?

Muchos evaluadores se impacientan por ver terminada la planeación de la evaluación, y por eso se apresuran en la recopilación de datos e intentan hacer trabajo exploratorio al mismo tiempo. Pero terminar un buen análisis del estado inicial es algo crítico para conocer una intervención y puede ahorrar tiempo y dinero en la evaluación, asegurar que la evaluación satisfaga las necesidades del cliente y sostener o iniciar relaciones, no sólo con el cliente, sino también con las partes interesadas clave. Más importante aún, un buen análisis del estado inicial puede asegurar que la evaluación plantee las preguntas correctas para obtener información que se necesita, en lugar de recopilar datos que nunca se utilicen.

Al principio de una evaluación, muchos evaluadores usualmente hacen supuestos, algunos de los cuales pueden ser incorrectos. Por ejemplo, pueden suponer que existe una rica infraestructura de datos cuando en realidad existe poca disponibilidad de datos. Suponen que consultores experimentados con extenso conocimiento del país les ayudarán con la evaluación, sólo para encontrarse con que las personas con las cuales contaban están ocupadas en otros proyectos. Se requiere un período exploratorio para conocer la disponibilidad de los datos y otros recursos.

También debe establecerse si es apropiada y posible la evaluación conjunta en la etapa del estado inicial. Si existe interés y es apropiado, los socios deberán determinar qué papel asumirá cada uno, y llegar a un acuerdo sobre temas como la oportunidad de la evaluación.

Equilibrar los costos y beneficios esperados de una evaluación

Durante la planeación del estado inicial en la agenda deben aparecer los costos y beneficios esperados de la evaluación y cómo equilibrarlos. Los beneficios de una evaluación incluyen los siguientes:

- Toma de decisiones basada en la evidencia que conduce a decisiones sensatas, en cuanto a qué ir aumentando gradualmente o replicar, qué mejorar o terminar, o qué ir disminuyendo de manera gradual.
- Contribución a flujos de conocimiento sobre lo que funciona bien (y en qué condiciones) y lo que no.
- Formación de capacidad local.

Los costos de las evaluaciones también son importantes, y deberían tenerse en cuenta en términos de:

- El costo del programa (no tiene sentido gastar US$50.000 para evaluar un programa de US$25.000).
- La carga de la evaluación en los beneficiarios del programa y otras personas a quienes se les solicita gastar tiempo suministrando información, o ayudando a los evaluadores en otra forma.
- Los costos de la reputación para el evaluador y la comunidad de evaluación, si es probable que se cuestionen los resultados debido a que la evaluación es de un programa altamente político o controvertido, o no se concede suficiente tiempo para efectuar una evaluación exhaustiva.

Deficiencias en el proceso de planeación del estado inicial

La creencia de que todo lo demás puede ocultarse –y de que si se efectúa la planeación del estado inicial, la evaluación procederá necesariamente en forma suave– es una deficiencia potencial del proceso de planeación del estado inicial. Otras deficiencias son las siguientes:

- Resistencia a modificar el plan original (Leeuw, 2003).
- La "macdonalización" de la sociedad, "el proceso por el cual se imponen los principios del restaurante de comidas rápidas en cada vez más sectores de la sociedad estadounidense, y también del resto del mundo" (Ritzer, 1993, p. 1) (esta frase es particularmente aplicable cuando las listas de verificación, listas de lo que hay que hacer y marcos de referencia reemplazan el pensamiento reflexivo).
- Creencias establecidas ("truismos") que surgen al efectuar la planeación del estado inicial (¿"Experimentos al azar? De ninguna manera: son demasiado complicados, demasiado costosos y demasiado difíciles en el contexto del desarrollo").
- "Pensamiento grupal" (estar de acuerdo con la posición del grupo para seguir siendo parte de él, a pesar de tener preocupaciones sobre la posición).
- Ponderación desproporcionada de las opiniones de los poderosos (ponderar automáticamente el valor de las sugerencias por el estatus de quienes las hacen).

Identificar el cliente principal y las partes interesadas clave

Una parte importante del análisis del estado inicial es identificar al cliente principal y las partes interesadas clave del proyecto, programa o política. No siempre pueden identificarse en forma directa las partes interesadas clave.

El cliente principal

Usualmente una parte interesada o un grupo de partes interesadas clave patrocina o solicita una evaluación y es el principal receptor de sus productos. Esta parte interesada o grupo de partes interesadas es el cliente principal de la evaluación. Las necesidades de este cliente tendrán una gran influencia en la evaluación.

El cliente principal puede:

- Autorizar y financiar el programa.
- Autorizar y financiar la evaluación.
- Ser responsable ante el público por la intervención.
- Ser la parte ante la cual son responsables los evaluadores.

Existe un solo cliente principal. Es importante reunirse con él (o con sus representantes, como una junta directiva, en el caso de un cliente grupal) al principio para identificar los temas de la evaluación desde su perspectiva. Durante esta reunión los evaluadores deben hacer preguntas sobre las necesidades del cliente en cuanto a la oportunidad y el uso previsto de la evaluación. El evaluador que primero escucha e indaga al cliente para determinar cuestiones subyacentes a la solicitud de evaluación, puede volver más tarde para presentar al cliente opciones amplias sobre las formas como puede enfocarse la evaluación, o recomendar un solo enfoque para la evaluación.

Partes interesadas

Partes interesadas: Persona u organización distinta del cliente que tiene algún interés en la intervención.

Las partes interesadas son las personas y organizaciones distintas del cliente que tienen algún interés en la intervención. Suelen ser aquellos que se ven afectados por la intervención, sea durante su realización o en años posteriores. Es importante incluir como partes interesadas a quienes usualmente no se les solicitaría participar en la evaluación.

Entre las partes interesadas están:

- Los participantes.
- Los beneficiarios directos.
- Los beneficiarios indirectos.
- Las organizaciones para el desarrollo que proveen financiación.
- Funcionarios del gobierno, funcionarios elegidos y empleados oficiales con intereses pertinentes, como planeadores y enfermeros de salud pública.

- Directores de programa, personal, miembros de juntas, administradores y voluntarios.
- Autores de políticas.
- La comunidad y grupos o asociaciones de interés, incluso aquellas que puedan tener una agenda distinta de los funcionarios del programa.

Las partes interesadas enfocan a menudo una intervención desde distintas perspectivas. El donante puede preocuparse porque se gaste el dinero en forma apropiada, y porque la intervención sea eficaz. El administrador del programa se preocupa por el buen manejo de la intervención, y porque genere buenas enseñanzas. Los participantes en el programa desean recibir más o mejores servicios. Los autores de políticas se preocupan porque la intervención cause el impacto deseado, y otros integrantes de la comunidad desean replicar o expandir la intervención, o limitar lo que perciben como consecuencias negativas de ella. Esta diversidad de intereses es algo bueno y puede revelarse en las discusiones iniciales.

Los papeles de cada individuo o grupo, con respecto a la evaluación y su uso potencial, deben identificarse claramente (cuadro 4.1).

Identificar e involucrar a partes interesadas clave

Puede identificarse a las partes interesadas clave, examinando los documentos sobre la intervención y conversando con el cliente principal de la evaluación, los patrocinadores del programa, el personal del mismo, los funcionarios locales y los participantes. Puede entrevistarse a las partes interesadas inicialmente, o reunirlos en pequeños grupos. Al contactar a las partes interesadas sobre la evaluación, el planeador debe ser claro en cuanto a cuál es el propósito al conversar con cada parte (advirtiéndoles sobre la evaluación que va a realizarse, y solicitándoles identificar los problemas que desearían se trataran en la evaluación).

Aumentar el uso de una evaluación es un proceso que empieza con una reunión con el cliente principal, y vinculando a las partes interesadas en el diseño de la evaluación. No es algo que suceda cuando está completo el informe de la evaluación y a punto de difundirse. Para algunas evaluaciones, las reuniones con las partes interesadas clave se efectúan periódicamente, o se establece una estructura más formal. El administrador de la evaluación puede configurar una estructura de comité asesor o de dirección.

Vincular a las partes interesadas desde un principio le da al evaluador un mejor conocimiento de la intervención, de lo que se intenta lograr y de los asuntos y problemas que se encontraron en su ejecución. El equipo de evaluación estará mejor informado sobre los asuntos que deben cubrirse en la evaluación y también de la información específica requerida. Las reuniones con las partes interesadas contribuyen a asegurar que la evaluación no omita asuntos críticos principales, y también ayuda a obtener "aceptación" de la evaluación: es probable que permitir que las partes interesadas sepan que los asuntos y preguntas que plantean se analizarán cuidadosamente, aumente su apoyo e interés en la evaluación. Así es como se construye el uso de la evaluación.

Cuadro 4.1 *Los papeles de las partes interesadas en una evaluación*

Parte interesada	*Hacer política*	*Tomar decisiones operativas*	*Proveer insumos a la evaluación*	*Reaccionar*	*Sólo por interés*
Desarrolladores del programa					
Patrocinadores financieros del programa					
Funcionario, junta o agencia que autoriza					
Proveedores de otros recursos (instalaciones, suministros, contribuciones en especie)					
Jefes y administradores superiores de agencias de implementación					
Administradores del programa responsables de la ejecución					
Personal del programa					
Personal de evaluación					
Beneficiarios directos del programa					
Beneficiarios indirectos del programa					
Adoptantes potenciales del programa					
Personas excluidas del programa (por ejemplo, por criterios de ingreso)					
Personas que reciben efectos negativos del programa o la evaluación					
Personas que pierden poder como resultado del programa					
Personas que pierden oportunidades como resultado del programa					
Integrantes de la comunidad, y el público					
Otros					

Fuente: los autores.

El grado en que las partes interesadas participen activamente en el diseño y la implementación de la evaluación depende de varios factores, incluyendo recursos y relaciones. Por ejemplo, puede ser que las partes interesadas dispongan de tiempo para distraerse de sus deberes regulares, o puede haber razones políticas para que la evaluación sea lo más independiente posible.

Efectuar un análisis de las partes interesadas

Para el análisis de las partes interesadas se han desarrollado muchas guías. En su sitio web, *A Guide to Managing for Quality,* el Management Sciences for Health Sciences y el Fondo de las Naciones Unidas para la Infancia (1998) disponen uno de esos procesos para identificar y evaluar la importancia de las personas, los grupos de personas o instituciones clave que pueden tener una influencia significativa en el éxito de una evaluación, y sugieren además estas razones para efectuar un análisis de las partes interesadas:

- Identificar personas, grupos e instituciones que pueden influir en la evaluación (positiva o negativamente).
- Anticipar la clase de influencia, positiva o negativa, que estos grupos tendrán en la evaluación.
- Desarrollar estrategias para obtener el apoyo más eficaz posible para la iniciativa, y reducir los obstáculos para el éxito de la implementación de la evaluación.

En el recuadro 4.1 se presenta un modelo para realizar un análisis de las partes interesadas. Aunque similar al cuadro 4.1, hace hincapié en generar apoyo para la evaluación y reducir la oposición a ella.

Si bien es muy importante ser incluyente, lo es también no serlo en exceso. Los esfuerzos para incluir a quienes se encuentran en la periferia podrían sólo irritarlos.

Algunas veces los evaluadores incluyen directamente una o más partes interesadas en la planeación y la ejecución de la evaluación. (Las evaluaciones participativas se discuten en el capítulo 5.) El evaluador facilita la participación de la parte interesada para:

- Formular los términos de referencia.
- Seleccionar el equipo de evaluación.
- Analizar los datos.
- Identificar hallazgos y formular conclusiones y recomendaciones (Mikkelsen, 2005).

Conocer el contexto

En un análisis del estado inicial se investiga también la relación entre las etapas del programa y el propósito general de la evaluación. La vida de un proyecto, programa o política puede tomarse como una progresión en la que se hacen distintas preguntas de evaluación en las diferentes etapas. Por ejemplo, no tendría sentido evaluar si se han

Recuadro 4.1 *Cómo efectuar un análisis de partes interesadas*

1. Realizar una reunión de intercambio de ideas con colegas para identificar personas, grupos e instituciones que afectarán o se verán afectados por la intervención, y relacionarlos en la columna de partes interesadas en el cuadro.
2. Identificar los intereses específicos que cada parte interesada tiene en la evaluación. Considerar temas como los beneficios potenciales de la evaluación para la parte interesada, los cambios en la evaluación que puede requerir la parte interesada que se hagan y las actividades del proyecto que puedan perjudicar o entrar en conflicto con la parte interesada. Registrar estos intereses en la columna "intereses de la parte interesada en el proyecto, programa o política".
3. Identificar la importancia de los intereses de la parte interesada para el éxito de la evaluación. Considerar: a) el papel que debe tener la parte interesada clave para que la evaluación sea exitosa y la posibilidad de que ésta represente ese papel, y b) la probabilidad y el impacto de una respuesta negativa de la parte interesada a la evaluación. Para cada parte interesada, se debe registrar la valoración en la columna "valoración de impacto potencial", asignando una "E" para extremamente importante, una "R" para razonablemente importante y una "N" para no muy importante.
4. Considerar las acciones que podrían ejercerse para obtener el apoyo de la parte interesada y reducir la oposición. Considerar cómo podría ser la aproximación a cada una de las partes interesadas. ¿Qué temas deseará la parte interesada que se traten en la evaluación? ¿Qué tan importante es incluirla en el proceso de planeación? ¿Existen otros grupos o individuos que puedan influir en la parte interesada para apoyar la evaluación? Registrar las estrategias para obtener el apoyo o reducir los obstáculos a la evaluación en la última columna del cuadro.

Cuadro del recuadro. *Modelo de ejemplo para efectuar el análisis de las partes interesadas*

Parte interesada	*Intereses en el proyecto, programa o política*	*Valoración del impacto potencial de la evaluación en la parte interesada y ésta en la evaluación*	*Estrategias potenciales para obtener apoyo o reducir obstáculos*

Fuente: Management Sciences for Health y Fondo de las Naciones Unidas para la Infancia, 1998.

cumplido los objetivos del programa sólo unos meses después de haber comprometido los fondos. En esta etapa temprana, una pregunta más apropiada podría ser la de si el programa ha conseguido los insumos necesarios para la ejecución. Pancer y Westhues (1989) presentan una tipología para esta progresión de las etapas del programa y las

preguntas de la evaluación (cuadro 4.2). Las preguntas que incluyen sólo son ejemplos y en cada etapa pueden hacerse muchas preguntas potenciales.

En el análisis del estado inicial otro paso es determinar el contexto de políticas. En la investigación se pueden identificar evaluaciones realizadas de programas similares. El evaluador empieza obteniendo informes de evaluaciones y revisándolos en cuanto a los temas tratados, el tipo de enfoque seleccionado, los instrumentos utilizados y los hallazgos. Si la evaluación es de una intervención nueva, puede ser necesario que se diseñe sin las guías de evaluaciones anteriores. Sin embargo, esto pocas veces sucede.

Cuadro 4.2 *Preguntas para hacer en las distintas etapas de una evaluación*

Etapa de desarrollo del programa	*Pregunta de la evaluación*
1. Valoración del problema y de las necesidades sociales	¿Hasta qué punto se satisfacen las necesidades de la comunidad?
2. Establecimiento de metas	¿Qué debe hacerse para satisfacerlas?
3. Diseño de alternativas del programa	¿Qué servicios podrían utilizarse para producir los cambios deseados?
4. Selección de una alternativa	¿Cuál de los posibles enfoques del programa es más fuerte?
5. Implementación del programa	¿Cómo debe empezar a operar el programa?
6. Operación del programa	¿Está operando el programa según se planeó?
7. Resultados/efectos/impacto del programa	¿Está causando el programa los efectos deseados?
8. Eficiencia del programa	¿Se obtienen los efectos del programa a un costo razonable?

Fuente: adaptado de Pancer y Westhues, 1989.

Acudir al conocimiento existente

En un análisis del estado inicial se investiga el conocimiento teórico y empírico existente sobre el proyecto, el programa o la política, lo que se conoce también como acudir al *fondo de conocimientos.*

El conocimiento proveniente de evaluaciones y otra investigación de las ciencias sociales, incluyendo los estudios económicos, aumenta cada día. Las revistas que contienen artículos que sintetizan la acumulación de conocimiento explicatorio sobre un tema específico; por ejemplo, el efecto del tamaño de la clase en el aprendizaje de los programas de nutrición para madres embarazadas en el peso del niño al nacer. La investigación orientada a problemas sobre el funcionamiento de las organizaciones combina teorías e investigación de disciplinas tan diversas como sociología organizativa, psicología cognitiva, teoría de elección pública

Fondo de conocimientos: Cuerpo de conocimiento teórico y empírico existente sobre el proyecto, el programa o la política.

y derecho y economía (Scott, 2003; Swedberg, 2003). Algunas organizaciones, como la Campbell Collaboration (http://www.campbellcollaboration.org), revisan la calidad de las evaluaciones sobre un tema determinado y sintetizan aquellas que satisfacen sus criterios. Los repositorios de experimentos al azar en el campo de la justicia criminal y la prevención del crimen, los programas de bienestar social y educativos y de salud indican que se han efectuado más de 10.000 "experimentos" (Petrosino y otros, 2003). Al organizar y planear una evaluación, es de crucial importancia identificar y revisar el fondo de conocimientos pertinente (recuadro 4.2).

Construir, utilizar y evaluar una teoría del cambio

La última parte del análisis del estado inicial es construir una teoría del cambio, y comprender cómo utilizarla y evaluarla. La lógica subyacente o la teoría del cambio es un tema importante para la evaluación, ya sea durante la etapa *ex ante* o *ex post* de un estudio. En esta sección se cuestiona sobre por qué utilizar una teoría del cambio y cómo construirla y evaluarla.

Aunque las definiciones varían (*ver* el capítulo 3), una de ellas de teoría del cambio afirma que es "una herramienta innovadora para diseñar y evaluar iniciativas de cambio social", una especie de "diseño de los bloques de construcción" necesarios para cumplir las metas a largo plazo de una iniciativa de cambio social (ActKnowledge and Aspen Institute, 2003). Una teoría del cambio puede igualmente considerarse como una representación de cómo se espera que una organización o una iniciativa obtenga resultados y una identificación de los supuestos subyacentes en ella.

Una teoría del cambio debe:

- Describir una secuencia de los insumos que utilizará el proyecto, programa o política; las actividades que apoyarán los insumos, los productos hacia los cuales el proyecto, el programa o la política está presupuestando (una actividad individual o una combinación de actividades) y los efectos directos y los impactos esperados.

Recuadro 4.2 *Acudir al fondo de conocimientos sobre la prevención del crimen*

El análisis de los datos sobre prevención del crimen sugiere que hubo 29 programas que funcionaron, 25 no y 28 promisorios (la información sobre otros 68 programas no estaba clara). Estos hallazgos se basan en una síntesis de más de 600 evaluaciones, incluyendo aquéllas de prevención de crimen basadas en el colegio y en la familia, programas de reducción de robos menores, arrestos por uso de estupefacientes, puntos de vigilancia o situación precaria, iniciativas de circuito cerrado, vigilantes comunitarios, programas de mentores y varias clases de sanciones y programas carcelarios (manejo de la ira, programas de capacitación, y cognitivos centrados en reducir la reincidencia, campamentos de entrenamiento, etcétera).

Fuente: Sherman y otros, 2002).

- Identificar casos o condiciones que puedan afectar la obtención de los efectos directos.
- Identificar los supuestos que se hacen en el programa sobre causas y efectos.
- Identificar supuestos críticos que la evaluación necesita examinar, basados en la política y el contexto ambiental y en una revisión de la literatura.

Es particularmente necesario identificar los casos o condiciones que puedan afectar la obtención de los efectos directos deseados de la evaluación, dada la reciprocidad y la complejidad de los programas de desarrollo. Las instituciones de desarrollo internacional ya proveen préstamos programáticos, lo que le da a los países en desarrollo una mayor discreción que la financiación de proyectos. Como indica Pawson (2006):

> Un cambio importante en las políticas públicas en los años recientes ha sido el aumento de programas complejos, de objetivos múltiples, en varias ubicaciones, de múltiples agencias y de múltiples objetos… Las razones son claras pues las raíces de los problemas sociales se entrecruzan. Un déficit en la salud puede originarse por la desventaja educativa, la desigualdad del mercado laboral, las disparidades ambientales, la exclusión en la vivienda, los patrones diferenciales de persecución del crimen, y otros. Quienes toman las decisiones han empezado, en consecuencia, a sopesar si las intervenciones de una sola medida o de un solo tema pueden estar tratando únicamente los síntomas.

Pawson piensa que en esos casos los evaluadores deberían:

- Comprender la teoría programática subyacente.
- Examinar la evidencia existente mediante la síntesis de la investigación.
- Considerar un programa complejo como cadenas de intervención, en las que un grupo de partes interesadas provee los recursos (materiales, sociales, cognitivos o emocionales) a otros grupos de parte interesadas, con la esperanza de que el cambio en el comportamiento vendrá luego.

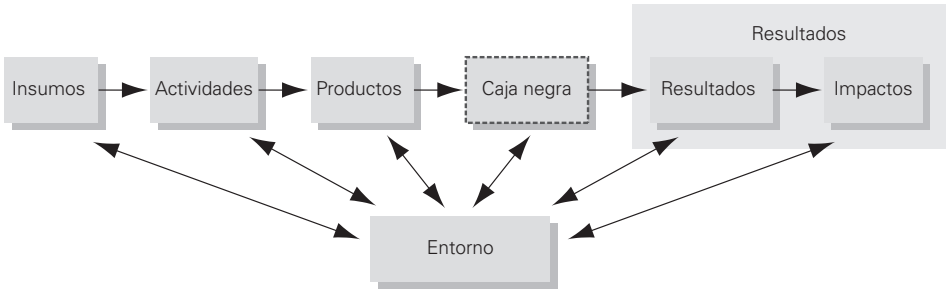
Cuando se planea una intervención, debe prestarse atención a la cuestión de cómo se construirá y probará la teoría del cambio subyacente al programa. Deben utilizarse efectos visuales para obtener una visión de conjunto de los componentes clave y las interacciones del proyecto, el programa o la política. Debería mostrar las causas y los efectos de los proyectos, los programas o las políticas, los eslabones de una cadena de razonamiento sobre qué es lo que causa qué cosa. El impacto o la meta deseada se presenta con mucha frecuencia como el último eslabón del modelo.

El valor de una teoría del cambio es que transmite visualmente las creencias de por qué es probable que el proyecto, el programa o la política tenga éxito en lograr sus objetivos. La teoría del cambio especifica asimismo los componentes de un programa y sus interrelaciones. Se proveen los recursos para que una organización realice sus actividades para alcanzar objetivos específicos. Estos recursos, actividades, productos, efectos directos deseados e impactos están interrelacionados.

En algunos casos, los evaluadores encuentran que una intervención ya tiene una teoría del cambio. De ser así, ellos deben revisarla cuidadosamente. En muchos casos, será necesario refinar o reelaborar la teoría del cambio existente y confirmarla con las personas directamente involucradas. Si no existe teoría del cambio, los evaluadores deberían crear una y validarla con el administrador y el personal del programa, de ser posible.

Con una teoría del cambio deben identificarse también los supuestos. Además, identificarse los supuestos más críticos que en la evaluación se deben someter a prueba (con base en el entorno predominante político y de políticas, y en la revisión de la literatura). Las teorías del cambio abren la "caja negra" para mostrar cómo espera una intervención convertir insumos, actividades y productos en resultados (gráfico 4.1).

Gráfico 4.1 *Paso de insumos a resultados*



Fuente: los autores.

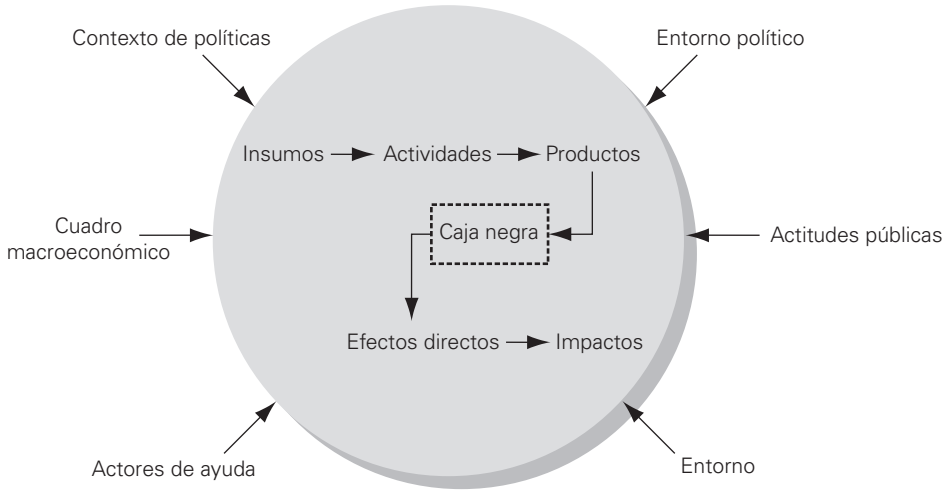
Es importante identificar qué sucede en el contexto general; esto es, el entorno en donde opera el programa. Este entorno (político, macroeconómico, de políticas, entre otros) influye en todas las partes del sistema (gráfico 4.2).

Detrás de todo proyecto, programa y política existe una teoría del cambio, que a menudo se espera sea articulada. La teoría puede representarse de manera visual en distintas formas, utilizando diferentes formatos o modelos. Estas representaciones se conocen algunas veces como *modelos teóricos, modelos lógicos, marcos de cambio, marcos lógicos, modelos de cadena de resultados* y *modelos de efectos directos.* Cada término es una variación sobre un tema que describe la teoría del cambio, y ésta debería disponer una cadena causal, mostrar influencias e identificar los supuestos básicos.

¿Por qué utilizar una teoría el cambio?

Una teoría del cambio es valiosa tanto para los evaluadores como para las partes interesadas, porque les permite trabajar juntos para construir "una visión de comprensión

Gráfico 4.2 *Influencias ambientales potenciales en los resultados del programa*



Fuente: los autores.

común de las metas a largo plazo, cómo se alcanzarán y qué se utilizará para medir el progreso por el camino" (ActKnowledge and Aspen Institute, 2003).

También puede utilizarse una teoría del cambio para informar sobre los resultados de una evaluación. En un informe de la Kellog Foundation (2004) se discute la importancia de la comunicación al informar sobre el éxito de un programa y su sostenibilidad. Identifica tres formas primarias como una descripción de una teoría del cambio puede apoyar esfuerzos estratégicos de mercadeo:

- Describir el programa en un lenguaje claro y lo suficientemente específico para ser comprendido y evaluado.
- Centrar la atención y los recursos en las operaciones prioritarias del programa y los resultados clave para propósitos de aprendizaje y mejoramiento del programa.
- Desarrollar comunicación y estrategias de mercadeo focalizadas.

En breve, articular la teoría del cambio para un proyecto, un programa o una política tiene varios beneficios:

- Contribuye a identificar elementos del programa que son críticos para su éxito.
- Contribuye a una comprensión común del programa y de las expectativas en las partes interesadas con base en un lenguaje común.
- Proporciona un fundamento para la evaluación.

- Identifica medidas para determinar el progreso en variables intermedias de las cuales dependen los efectos directos.
- Identifica los supuestos hechos, lo que puede convertirse en una base para investigar cuestiones de evaluación.

Construir una teoría del cambio

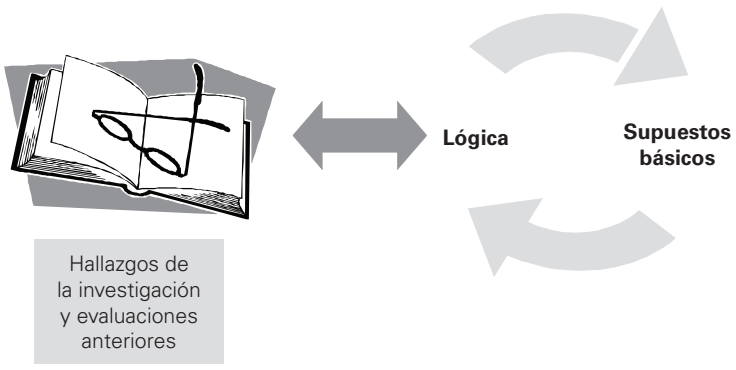
Los administradores desarrollan a menudo una teoría del cambio al conceptuar un proyecto, un programa o una política. Durante este proceso pueden incluir a las partes interesadas. Sin embargo, la teoría del cambio no siempre se hace explícita, ni es necesariamente consistente de principio a fin para una intervención dada. Para otras intervenciones, puede no existir una teoría del cambio. En este caso, el evaluador deberá construir una. La base de toda evaluación debería ser el examen de la teoría del cambio.

Antes de empezar a revisar o a construir una teoría del cambio, los evaluadores deben tener un conocimiento claro del propósito y las metas del proyecto, el programa o la política, y poner en consideración tres preguntas principales:

- ¿Se fundamenta el proyecto, el programa o la política en la investigación y la evaluación?
- ¿Cuál es la cadena lógica o de resultados del proyecto, el programa o la política?
- ¿Cuáles son los supuestos básicos que se hacen?

El proceso empieza con el mayor aprendizaje posible sobre intervenciones y evaluaciones relacionadas. Con la nueva información se inicia el proceso de prolongar la lógica del programa y los supuestos básicos. Al identificarse la lógica, se coloca en una cadena de eventos y se mapea o dibuja. Luego se identifican los supuestos básicos (gráfico 4.3).

Gráfico 4.3 *Proceso de construir una teoría del cambio*



Fuente: los autores.

¿Se fundamenta el programa en la investigación y evaluaciones anteriores?

El desarrollo de una teoría del cambio empieza identificando y revisando la literatura sobre investigación y evaluación anterior. Por ejemplo, la investigación anterior puede revelar que cuando se controlan otros factores, se encuentra una relación positiva entre el desempeño académico estudiantil y la participación paterna en el trabajo de casa del estudiante. O en una evaluación de un programa educativo anterior, en una zona urbana de un país en desarrollo, pueden haberse reportado hallazgos moderadamente exitosos para un componente de programa que incluye visitas del maestro al hogar de los estudiantes. Una vez se ha completado la investigación de la literatura y se han identificado hallazgos aplicables, puede construirse una teoría del cambio para predecir qué ocurrirá como resultado de una intervención similar. Puede ser necesario construir teorías del cambio sin el beneficio de los hallazgos de investigaciones o evaluaciones anteriores. En este caso, las teorías serán necesariamente más débiles.

Una revisión de la literatura de evaluación debería empezar con una identificación general de posibles fuentes de información sobre investigación y evaluación. Las fuentes incluyen las siguientes:

- Estudios de evaluación efectuados en la propia organización.
- El repositorio de publicaciones y documentos de la Organización para la Cooperación y el Desarrollo Económicos/Comité de Asistencia para el Desarrollo (OCDE/CAD), o la información por países.
- Estudios de evaluación efectuados por organizaciones para el desarrollo, bancos de desarrollo, ONG y otras.
- Artículos en revistas de evaluación.
- Investigación aplicada reportada en revistas de psicología, sociología, educación y otros campos sobre temas específicos.
- Investigación sobre teorías de desarrollo.

Los resúmenes ejecutivos y las conclusiones o enseñanzas aprendidas son un buen lugar para determinar la aplicabilidad de los informes de evaluación. Para artículos de investigación, los resúmenes pueden leerse rápidamente. Con limitación de tiempo, los evaluadores deberían explorar para hallar información importante. Una vez los evaluadores ubican hallazgos de investigación y evaluación pertinentes, deben examinarlos cuidadosamente (recuadro 4.3).

¿Cuál es la lógica del programa?

La lógica de un programa, una política o un proyecto se fija en el propósito y las metas de una intervención, y postula "Si sucede X, entonces debería suceder Y". Se especifican los detalles sobre la naturaleza del cambio deseado: la población objetivo, el nivel de cambio requerido para señalar el éxito, el marco de tiempo sobre el cual se espera ocurra dicho cambio. Se crea entonces una cadena de enunciados del tipo "si-entonces": la

Recuadro 4.3 *Revisar la literatura sobre programas para mejorar el logro estudiantil en un país de África meridional*

En una organización para el desarrollo, a los evaluadores se les solicitó evaluar un programa educativo en África cuya meta era mejorar el logro de los estudiantes en clase. A éstos se les informó que las visitas al hogar por parte de los maestros de primaria en un país vecino se asociaron con un mayor logro estudiantil. Los evaluadores empezaron explorando la literatura sobre investigación y evaluación y buscaron, primero, investigación sobre educación primaria y logros para ver qué hallazgos había disponibles sobre intervenciones similares. Se centraron en el *International Journal of Educational Research, la American Education Research Journal, Early Childhood Research and Practice* y el *Journal of Educational Research.* Además revisaron la Development Gateway, el CAD Evaluation Resource Centre y el Education Resources Information Center en búsqueda de evaluaciones. Mediante este proceso, los evaluadores encontraron alguna investigación relacionada y varios proyectos y programas que parecían ser similares. El equipo de evaluación leyó la literatura de investigación y evaluación a fin de aprender cuanto fuese posible sobre los asuntos y hallazgos de programas similares en otros países.

Fuente: los autores.

teoría del cambio. Pequeñas piezas de papel o tarjetas pueden utilizarse para registrar los enunciados en la cadena de actividades y se pueden mover, agregar o eliminar al construirse la cadena.

Los evaluadores trabajan a menudo hacia atrás, desde la meta a largo plazo, identificando la cadena lógica de casos, hasta regresar a la situación actual. Si se ha construido ya una teoría del cambio, el evaluador debe utilizar un proceso similar para reconstruirla.

Considérese el ejemplo de una intervención para capacitar gente en técnicas de evaluación del desarrollo (gráfico 4.4). Los resultados esperados de esta intervención serían producir evaluaciones de más alta calidad y la toma de mejores decisiones basadas en la evidencia. (La meta última serían las mejoras en el desarrollo, pero como son muchos los factores que influyen en éste, los evaluadores no intentan medir este efecto directo final.) Una simple cadena para esta intervención incluye lo siguiente: si los evaluadores están mejor capacitados, realizarán mejores evaluaciones, lo que debe producir información útil para los autores de políticas, lo que a su vez llevará a una toma de decisiones basada en la evidencia. La información útil debería producir mejores decisiones en quienes las toman.

Lo que se ha presentado hasta el momento es un modelo lineal. Una buena teoría del cambio no asume simples relaciones lineales de causa y efecto; muestra relaciones complejas utilizando cajas y flechas que enlazan hacia atrás a partes anteriores, o hacia delante a partes posteriores, de la teoría del cambio. También detallan los supuestos básicos que subyacen en el modelo, incluyendo los factores principales contextuales o ambientales, o los casos que probablemente influirán en la intervención.

001El camino.indd 144 08/11/2010 03:27:46 p.m.

Gráfico 4.4 *Una teoría del cambio simple para mejorar la toma de decisiones capacitando evaluadores*

Fuente: los autores.

¿Qué son los supuestos básicos?

Las cadenas de lógica iniciales aparecen mucho como lineales. Cuando los evaluadores analizan los muchos factores que interactúan con sus proyectos, programas y políticas, la teoría del cambio se hace más compleja. Al identificar la lógica del programa, deben identificar también los supuestos que hacen sobre el proceso del cambio. Los supuestos que constituyen el mayor riesgo para el éxito de la intervención (los supuestos básicos) pueden entonces analizarse y probarse como parte de la evaluación.

Los supuestos pertenecen usualmente a uno de estos cuatro grupos:

- Afirmaciones sobre las conexiones entre los efectos directos a largo plazo, intermedios y a corto plazo en el mapa.
- Sustanciación para la pretensión de que se han identificado todas las condiciones previas importantes para el éxito.
- Justificación de los vínculos entre las actividades del programa y los efectos directos que se espera produzcan.
- Conocimiento implícito o explícito de los factores contextuales o ambientales que probablemente apoyarán u obstaculizarán el progreso en obtener resultados y el grado en que se obtienen.

Los evaluadores estudian la lógica emergente e investigan sus supuestos. Las posibles preguntas incluyen las siguientes:

- ¿Es verosímil esta teoría del cambio? ¿Es probable que la cadena de eventos lleve a la meta a largo plazo?
- ¿Es factible esta teoría del cambio? ¿Es posible que las capacidades y los recursos para implementar las estrategias produzcan los efectos directos?
- ¿Es verificable esta teoría del cambio? ¿Se han especificado las mediciones para determinar el éxito? (Anderson, 2004).

También pueden preguntar:

- ¿Qué más está sucediendo en el entorno que pudiese ayudar u obstaculizar la intervención?

Los supuestos se escriben y se incluyen entonces en la cadena de casos. Pueden utilizarse pequeñas tarjetas, de modo que se reorganicen para que concuerden con la teoría emergente.

No todos los supuestos deben identificarse; si así se hiciera, la lista sería muy larga. Sin embargo, los supuestos básicos –los que presentan el mayor riesgo para el éxito del programa si se encuentra que son falsos– deben identificarse.

En el ejemplo del programa de capacitación, los supuestos básicos incluyen los siguientes:

- Los evaluadores no tienen fuentes fácilmente disponibles de capacitación que satisfagan sus necesidades.
- Los evaluadores pueden obtener los recursos financieros para participar en la capacitación.
- La capacitación es apropiada para las necesidades de los evaluadores.
- Los evaluadores valoran la capacitación y tienen motivación para aprender.
- Los evaluadores recibirán apoyo y otros recursos necesarios, de modo que puedan poner en práctica lo que aprenden en la capacitación.
- Los evaluadores adquirirán destrezas para elaborar informes, necesarios para comunicar la información eficazmente a la agencia del gobierno.
- Los funcionarios oficiales que toman las decisiones utilizarán los resultados de las evaluaciones para tomar mejores decisiones basadas en la evidencia.

Para que esta cadena sea eficaz, deben atenderse los supuestos críticos, que pueden listarse o dibujarse junto o dentro del diagrama de la teoría del cambio (gráfico 4.5).

Modelo para la teoría del cambio

La Kellog Foundation (2004) sugiere que los evaluadores utilicen un modelo para ayudarse en la explicación de la teoría del cambio (gráfico 4.6).

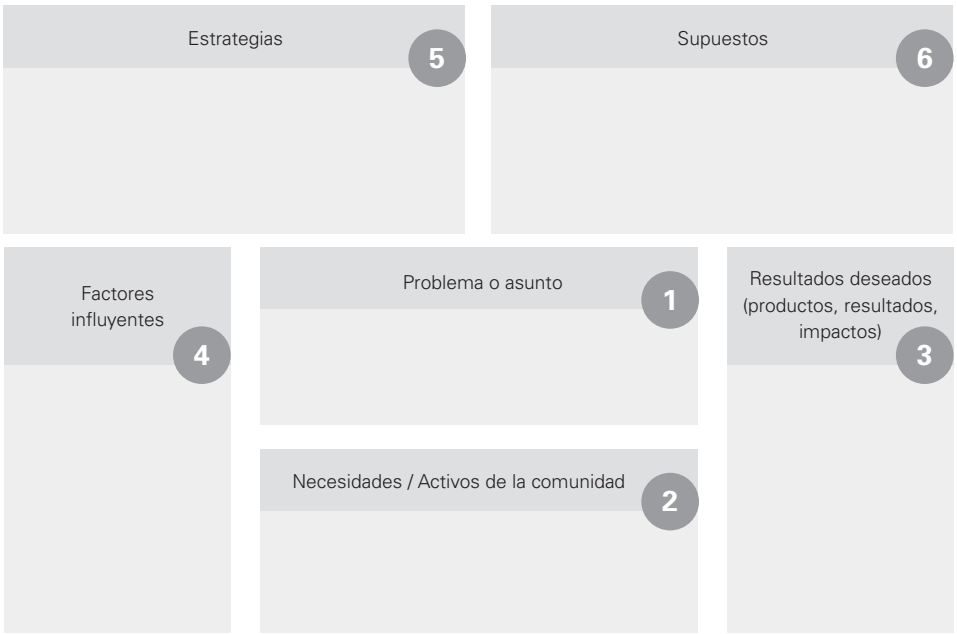
Para utilizar este modelo, la Kellog Foundation sugiere que los evaluadores empiecen por la mitad de éste (*problema o asunto*), la cual es el corazón del modelo y el centro de la teoría del cambio. En este espacio, el evaluador escribe un enunciado claro del problema que va a atender la intervención.

En el segundo paso, *necesidades/activos de la comunidad,* el evaluador especifica las necesidades y los activos de la comunidad u organización. Si se han valorado las necesidades o éstas han recibido prioridad, debe incluirse aquí esa información.

Gráfico 4.5 *Diagrama simple de la teoría del cambio con supuestos básicos identificados*

Fuente: los autores.

Gráfico 4.6 *Modelo para la teoría del cambio*



Fuente: Kellog Foundation, 2004.

En el tercer paso, *resultados deseados,* Kellog sugiere que el evaluador identifique qué se espera lograr con la intervención en el corto y el largo plazo. Esto puede ser principalmente una visión del futuro. Las entradas en este recuadro serán los productos, efectos directos e impactos.

En el cuarto paso, *factores influyentes,* Kellog le pide al evaluador listar las barreras y los apoyos potenciales que puedan afectar el cambio deseado. Éstos pueden ser factores de riesgo, entorno existente de políticas u otros factores. (Pueden provenir de la revisión de la literatura de investigación y evaluación o de otras fuentes.)

En el quinto paso, *estrategias,* el evaluador debe listar las estrategias exitosas generales que la investigación ha identificado, que ayudaron a comunidades u organizaciones similares a lograr la clase de resultados que el proyecto, el programa o la política intenta producir ("buenas prácticas").

En el último paso, *supuestos,* Kellog le pide al evaluador enunciar los supuestos sobre cómo y por qué las estrategias de cambio identificadas funcionarán en la comunidad u organización. Éstos pueden ser principios, creencias o ideas. El modelo para la teoría del cambio puede utilizarse entonces para prolongar la representación gráfica de la teoría del cambio.

Si la construcción de la teoría del cambio es tarea de un grupo de personas, cada persona (o grupo) puede recibir un modelo en blanco para llenarlo. Cuando todos los modelos se encuentren llenos, el grupo puede discutirlos y llegar a un acuerdo.

Este es un ejemplo de modelo. Podría revisarse o ajustarse para satisfacer necesidades o prácticas organizativas específicas. Por ejemplo, un evaluador podría desear indicar no sólo estrategias exitosas sino también no exitosas (para evitarlas).

Ejemplos de construcción de teoría del cambio

Con dos ejemplos se ilustra cómo un evaluador construye una teoría del cambio. En el primero se describe un programa para mejorar el rendimiento académico, haciendo que los maestros visiten los hogares de los estudiantes. En el segundo se describe un programa en que se utilizan talleres de participación para reducir la corrupción en el gobierno.

Mejorar el rendimiento académico mediante visitas de los maestros a los hogares de los estudiantes. Considérese la situación siguiente: debe hacerse la evaluación a un programa maduro, pero éste no tiene una teoría del cambio. El evaluador debe empezar desarrollándola y acudir a la literatura sobre investigación y evaluación (según se describió en el recuadro 4.3). La meta deseada del programa es mejorar el rendimiento académico de los estudiantes y la intervención consiste en que los maestros visiten a los estudiantes en sus hogares. La lógica de la situación es ésta: si los maestros (insumo) visitan (actividad) los hogares de sus estudiantes (insumo) y charlan con los padres (producto) tendrán mayor afinidad con el estudiante (efecto directo). Los padres comprenderán mejor la necesidad del colegio de que los estudiantes completen sus tareas a tiempo y de que asistan todos los días al colegio (producto); como resultado, se asegurarán de que sucedan

ambas cosas (efecto directo). Como el estudiante hace sus tareas, asiste regularmente al colegio y tiene afinidad con su maestro, mejorará su rendimiento (impacto).

El evaluador que desarrolla la teoría del cambio empieza con el resultado deseado –mejores logros en lectura–, y lo ubica en la parte inferior del diagrama, en este caso ver gráfico 4.7. A continuación, el evaluador identifica la intervención escribiendo "visitas de maestros al hogar del estudiante" en la parte superior. De allí, el evaluador identifica tres resultados posibles de las visitas:

- Los maestros obtienen un conocimiento de la cultura del hogar del estudiante.
- Los padres obtienen información sobre lo que el colegio espera del estudiante.
- Los maestros y padres pueden abordar los problemas que impiden al estudiante tener un buen rendimiento en el colegio.

De cada uno de estos tres posibles resultados, el evaluador identificará otros, creando cadenas de resultados e interactuando éstos. Por ejemplo, del resultado "conocimiento del maestro de la cultura del hogar del estudiante", el evaluador identifica una cadena de resultados que incluye lo siguiente:

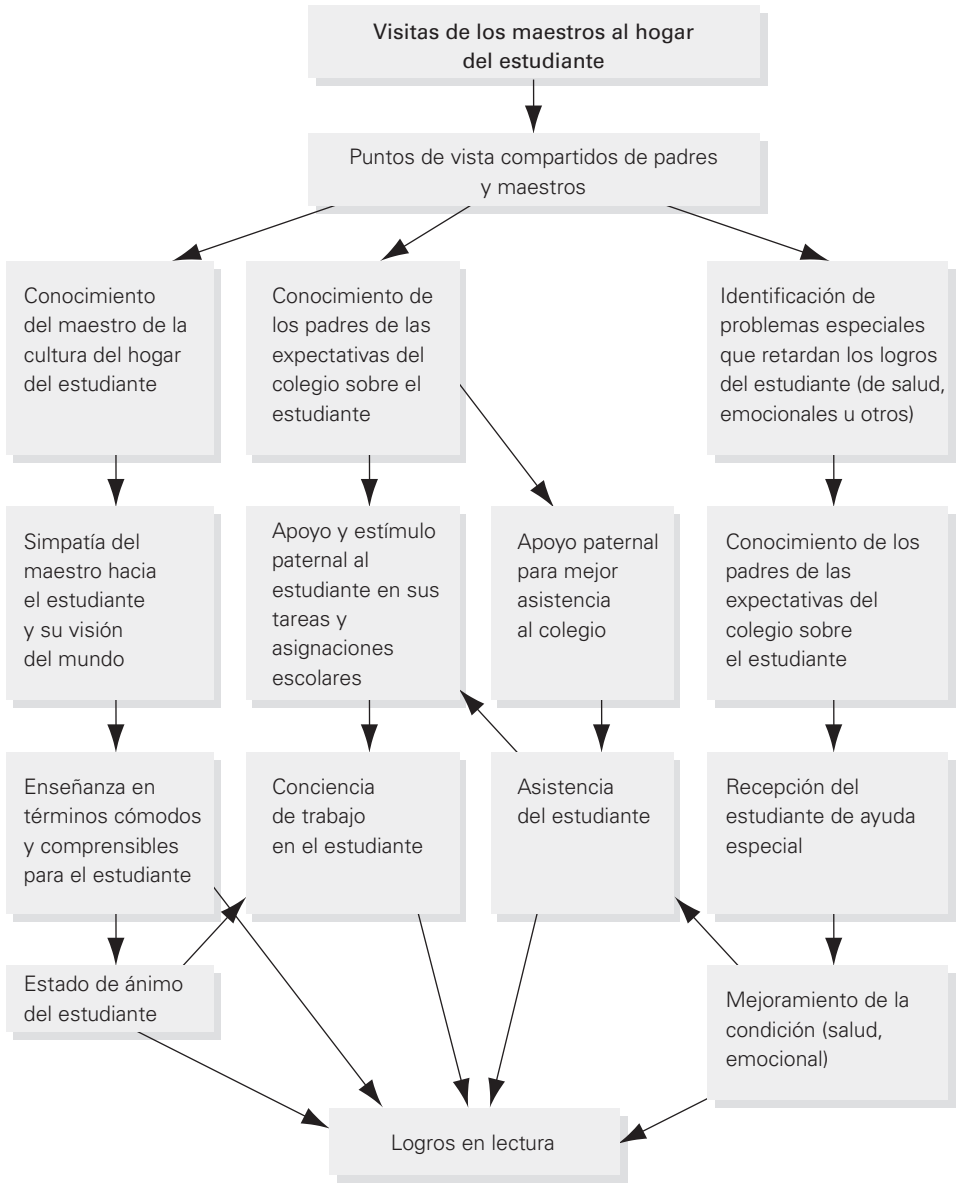
- Entonces los maestros tienen simpatía por sus estudiantes y su visión del mundo.
- Entonces los maestros empiezan a enseñar en formas más cómodas para el estudiante.
- Entonces mejora el estado de ánimo del estudiante.
- Entonces progresa el logro del estudiante en lectura.

El evaluador identifica luego otros resultados posibles de cada una de las posibilidades originales, los cuales terminan todos con el progreso en lectura. Algunas cadenas de resultados interactúan también con los demás resultados posibles.

Con esta teoría del cambio se identifican también varios supuestos. En este caso, se listan en lugar de incluirse en el diagrama. Los supuestos que el evaluador identifica son:

- Los estudiantes provienen de familias con padre y madre, con casa.
- Los padres están disponibles en la casa cuando los maestros van a ella.
- Los maestros están dispuestos a hacer las visitas.
- Los padres reciben bien a los maestros en su casa.
- Los padres se sienten cómodos discutiendo sus opiniones sobre la educación de su hijo.
- Los maestros conocerán mejor la cultura del hogar y así tendrán mayor simpatía con el estudiante.
- Los maestros ajustarán su enseñanza y estilo de enseñar utilizando lo que observen en las visitas.
- Los padres desearán participar en el aprendizaje del estudiante, y desearán que su hijo asista al colegio, haga sus tareas y obtenga logros académicos.
- Los padres no necesitan que su hijo trabaje.
- No sucede nada más en el colegio o en el entorno del hogar que pueda causar mejoras en el logro del estudiante.

Gráfico 4.7 *Teoría del cambio para un programa de mejoramiento del rendimiento estudiantil en lectura, mediante visitas de los maestros al hogar del estudiante*



Fuente: Weiss, 1972, 50.

La evaluación se desarrollaría con base en los supuestos que la literatura o las partes interesadas han identificado como críticos.

Utiliza talleres de participación para combatir la corrupción. Considérese un ejemplo diferente: un programa intenta introducir talleres de participación para reducir la corrupción gubernamental. Para desarrollar la teoría del cambio, el evaluador empieza escribiendo la meta a largo plazo de reducir la corrupción en la parte inferior del diagrama, y escribiendo la intervención en la parte superior del mismo. Los principales eventos que se predice ocurrirán se colocan en una cadena de eventos entre los dos (el objetivo y la intervención) (gráfico 4.8) (Haaruis, 2005).

Aunque se han identificado muchos supuestos, el evaluador limita los que se investigarán en la evaluación a tres básicos, basados en la revisión de la literatura y en las discusiones con las partes interesadas y clientes:

- Los talleres de participación son eficaces y satisfacen las necesidades de los participantes y del programa.
- Los participantes tienen destrezas, actitudes y motivaciones para participar en los talleres.
- Los participantes desarrollarán un sentido de la "propiedad local", creando un efecto de goteo.

Terminología y representaciones gráficas de la teoría del cambio

Por cuanto la teoría del cambio se ha desarrollado y llegado a constituir una fuerza importante en la evaluación, ha surgido alguna confusión sobre la terminología, que incluye términos como modelos lógicos, modelos de efectos directos y modelos teóricos. Por ejemplo, Patton (2002), distingue un modelo lógico de una teoría del cambio, afirmando que el único criterio para un modelo lógico es que presenta un orden razonable, defendible y secuencial, partiendo de los insumos, pasando por las actividades y terminando con los productos, los efectos directos y los impactos. En contraste, una teoría del cambio debe especificar y explicar también los vínculos causales que se asumen, sobre los que se formulan hipótesis y se ponen a prueba. Al margen de la terminología o el formato específico utilizado, todas las descripciones de la teoría del cambio deben diseñar una cadena causal, mostrar influencias e identificar supuestos básicos.

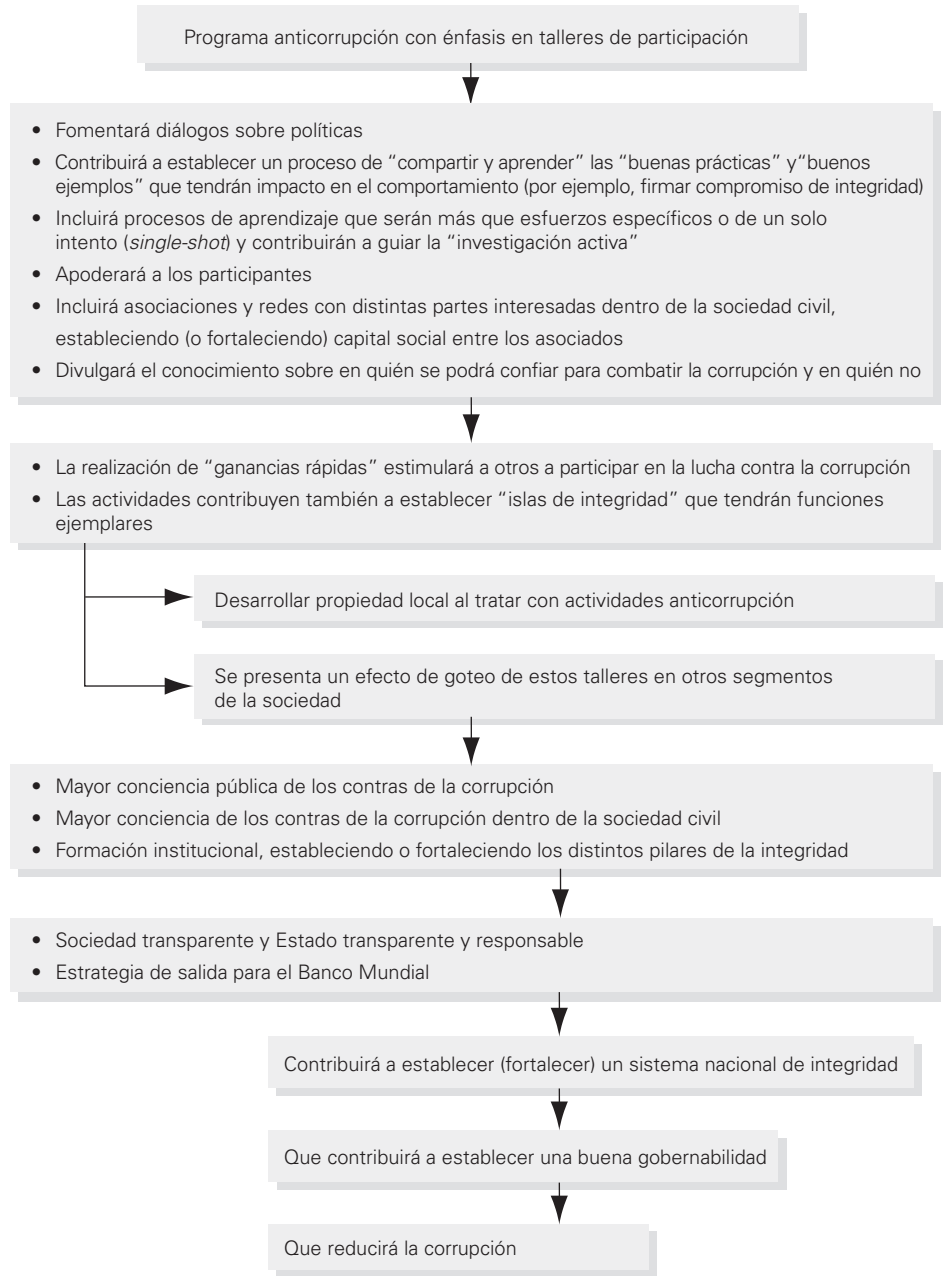
Los modelos de teoría del cambio pueden presentarse de manera visual en diagramas de flujo en distintas formas. En esta sección se ilustran algunos formatos.

Diagrama de flujo estándar

Diagrama de flujo: Diagrama que ilustra la secuencia de resultados producidos por las actividades y productos.

Los diagramas de flujo son el formato más común utilizado para ilustrar la teoría del cambio. Un *diagrama de flujo* ilustra la secuencia de resultados que fluyen (o resultan)

Gráfico 4.8 *Teoría del cambio para un programa de uso de talleres de participación para reducir la corrupción en el gobierno*



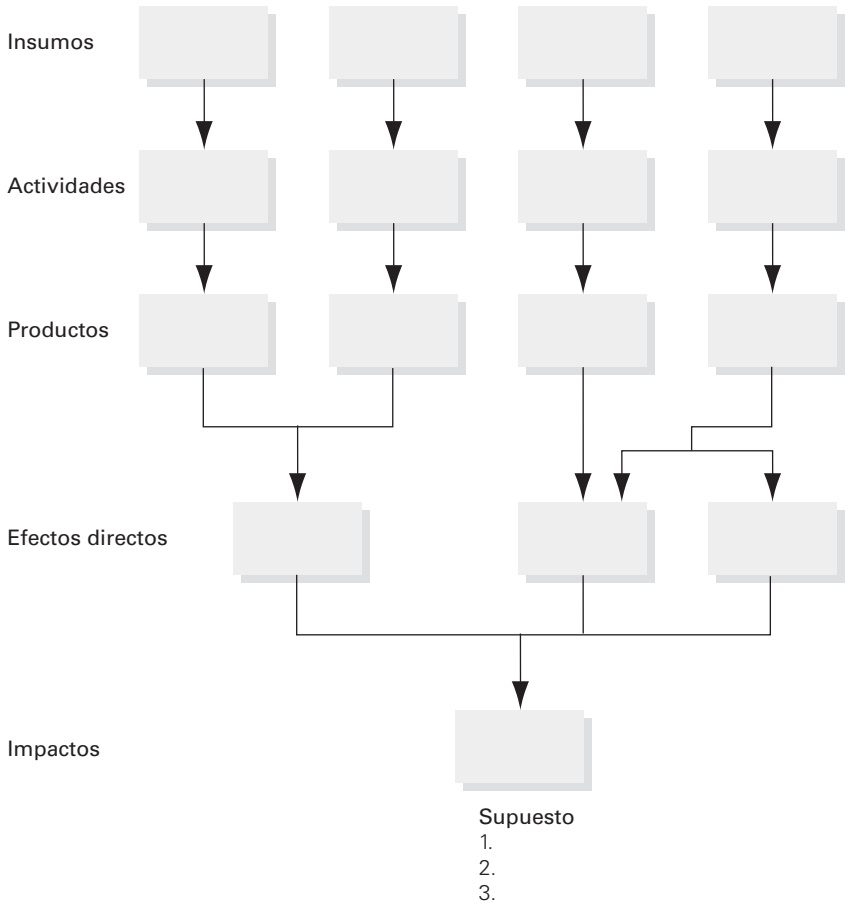
Fuente: Leeuw, Gils y Kreft, 1999.

de las actividades y los productos (gráfico 4.9). Es un modelo de lógica muy flexible en tanto que se presenten los componentes centrales: insumos, actividades, productos y resultados.

Pueden presentarse diferentes niveles de resultados (inmediatos, intermedios y finales) para garantizar que el diagrama de flujo indique cómo se obtienen los resultados finales. Cuando se utiliza este formato, los evaluadores deben listar sus supuestos, incluyendo los factores del entorno externo que puedan afectar la medida en que la intervención logre los resultados deseados.

Los vínculos de causa-efecto pueden explicarse utilizando enunciados "si-entonces". Por ejemplo, "si se implementa la actividad, se producirán estos productos. Si se obtienen estos productos, llevarán al primer nivel de resultados inmediatos".

Gráfico 4.9 *Diagrama de flujo estándar*



Fuente: los autores.

Diagrama de flujo estándar con énfasis en los supuestos

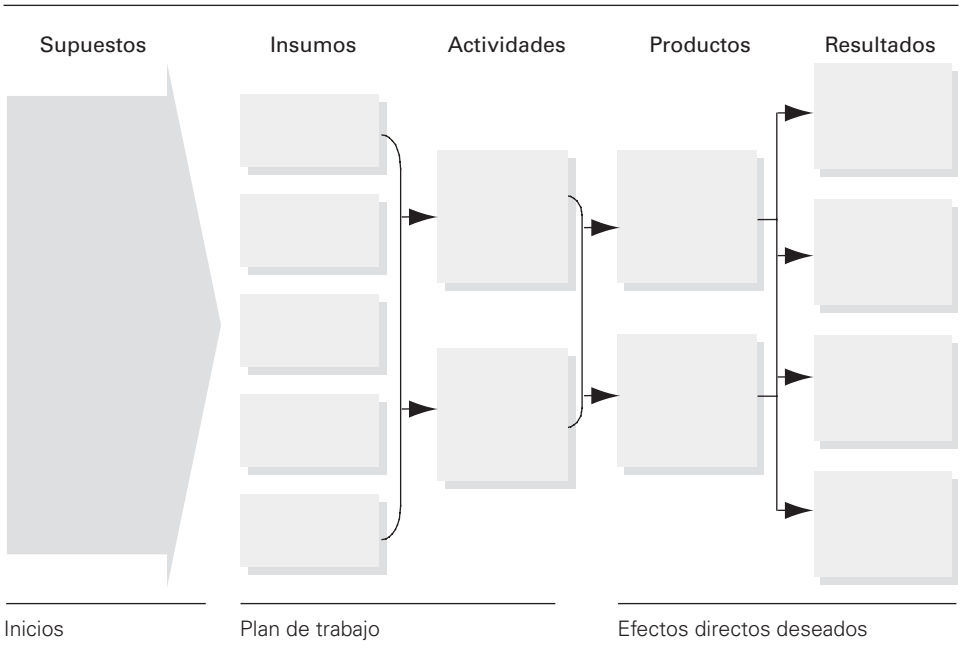
Otro diseño para una teoría del cambio se presenta en el gráfico 4.10. Este modelo incluye los supuestos como principios subyacentes en el diseño de la iniciativa.

Cadena de resultados estándar

Una *cadena de resultados* –conocida también como cadena de desempeño– es similar a un diagrama de flujo, pero no necesariamente empareja actividades específicas con productos o resultados específicos. Dado que no presenta el mismo detalle que un diagrama de flujo con respecto a la secuencia causal de actividades, productos y resultados, el usuario debe verificar que lo "simplista" no reemplace lo "estándar". A semejanza de otras representaciones visuales, la influencia de factores externos se considera en forma explícita. Los insumos, actividades y productos se utilizan a menudo como medidas de eficiencia, y los resultados se usan para establecer la eficacia del programa (gráfico 4.11).

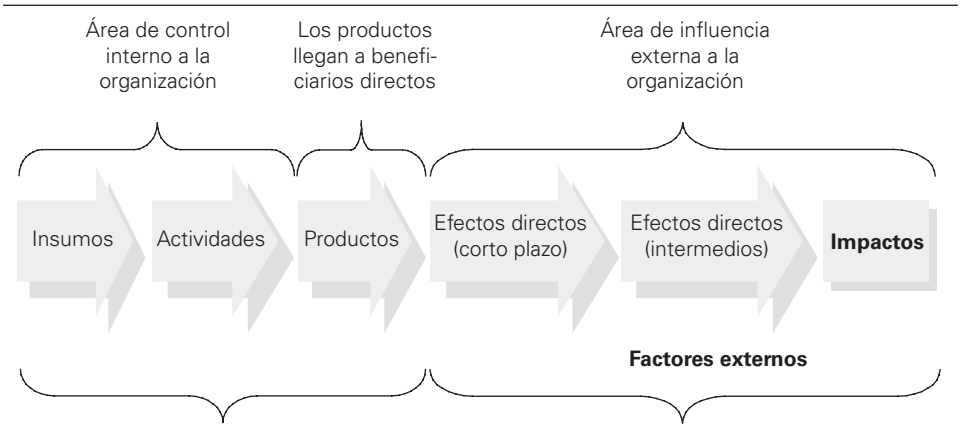
Cadena de resultados: Cadena causal para una intervención de desarrollo que estipula la secuencia necesaria para lograr los objetivos deseados, empezando con los insumos, pasando por las actividades y productos, y terminando con los resultados, los impactos y la retroalimentación.

Gráfico 4.10 *Diagrama de flujo con énfasis en los supuestos*



Fuente: adaptado de Kellog Foundation, 2004.

Gráfico 4.11 *Cadena de resultados estándar*



Fuente: Canadian Department of Foreign Affairs and International Trade, 2005.

Muchos ejemplos de modelos completos están disponibles en el sitio web de la University of Winsconsin Extension. Los enlaces al sitio aparecen al final de este capítulo.

Marco lógico (logframe)

Marco lógico (*logframe*): Matriz que vincula las actividades, los resultados, propósito y objetivos de un programa, una política o un proyecto, en una jerarquía.

Una variante del modelo de la teoría del cambio es el marco lógico, o como se le denomina comúnmente, el *logframe*. Un *logframe* vincula las actividades, los resultados, el propósito y los objetivos de un programa, una política o un proyecto en una jerarquía. Para cada objetivo del programa, el evaluador identifica los indicadores de desempeño que se necesitan, las fuentes de información para cada indicador y los supuestos. Los *logframes* clarifican los objetivos de un proyecto, programa o política, e identifican los vínculos causales entre insumos, procesos, productos, efectos directos e impacto. Muchas organizaciones de desarrollo requieren su uso y han capacitado a su personal para utilizarlos.

El *logframe* es esencialmente una matriz de 4 x 4 que resume los elementos críticos de un proyecto, programa o política. El enfoque trata las cuestiones clave en forma metódica, utilizando lógica causal.

El *logframe* puede utilizarse para estos propósitos:

- Mejorar la calidad de diseño de un proyecto, un programa o una política, requiriendo la especificación de objetivos claros, el uso de indicadores de desempeño y la valoración de riesgos.
- Resumir el diseño de actividades complejas.

- Asistir al personal en la preparación de planes operativos detallados.
- Proveer una base objetiva para revisar las actividades, el control y la evaluación (Banco Mundial, 2004).

El marco lógico ha sido criticado con atención, y las críticas importantes incluyen:

- Al desarrollar un marco lógico, un evaluador puede perderse en los detalles y perder de vista el cuadro general.
- No se hace énfasis en los datos básicos.
- Los marcos lógicos son demasiado simples, aun para diseños de proyecto simples. Como anota Gasper (1997), "no puede capturarse todo lo importante en un diagrama de una a tres páginas de cuatro o cinco niveles".
- Muchos usuarios de marco lógico omiten reconocer que un marco incluye algunas cosas y deja por fuera otras. El propósito de un marco es servir de ayuda para realizar una evaluación y no sustituirla.
- El marco lógico no tiene en cuenta los efectos directos imprevistos; su alcance se limita a los objetivos declarados.
- Después de haberse preparado un marco lógico, tiende a quedar fijo y no actualizarse, convirtiéndose así en un *lock-frame* ("marco enllavado") (Gasper, 1997).

Un marco lógico para una clínica infantil podría incluir las vacunaciones como una de sus actividades, con un resultado objetivo de vacunar un 50% de todos los niños menores de seis años en un distrito particular (cuadro 4.3). Si se lograra este objetivo, debería decrecer la incidencia de enfermedades infantiles prevenibles, y con este decrecimiento debería lograrse el objetivo general de reducir el número de muertes de niños menores de seis años.

En la segunda columna del cuadro 4.3 se identifican los indicadores que verifican el grado en que se ha logrado cada objetivo. En la tercera y cuarta columnas se especifica dónde se obtendrán los datos para evaluar el desempeño frente a los indicadores y los supuestos sobre la naturaleza y la accesibilidad de esos datos.

Evaluar una teoría del cambio

Una vez se ha desarrollado la teoría del cambio, los evaluadores deben retroceder y evaluar la calidad de la teoría desde distintos puntos de vista. Estos puntos de vista o marcos incluyen la evaluación:

- En relación con las necesidades sociales.
- De la lógica y la viabilidad.
- Mediante la comparación con la investigación y la práctica.
- Mediante la comparación de la teoría programática del cambio con una o más teorías científicas aplicables.
- Mediante la observación preliminar (adaptado de Rossi, Freeman y Lipsey, 1999).

001El camino.indd 157

Cuadro 4.3 *Marco lógico para el programa de reducción de la mortalidad infantil*

Componente	*Indicadores de desempeño*	*Verificación*	*Supuestos*
Meta: mejorar el bienestar económico y social de las mujeres y su familia	Mejoras en el ingreso familiar por porcentaje de familias participantes Mejoras en las medidas de condición de la salud, la nutrición y la participación educativa	Encuestas de hogares sobre condiciones económicas, sociales y de salud de todos los miembros de la familia	
Objetivo: ofrecer oportunidades a las mujeres para obtener ingresos y aprender mientras sus hijos están atendidos en centros diurnos de cuidado infantil	• Hogares de atención diurna en funcionamiento, con servicios de cuidados accesibles y económicos de calidad adecuada durante las horas laborales, permitiendo cambios en el empleo y las actividades educativas de la mujer	De las encuestas: cambios en el empleo y educación de las mujeres y su evaluación de los servicios prestados Evaluaciones de la calidad de los servicios prestados con base en la observación	Otros miembros de la familia mantienen o mejoran su empleo e ingresos Las condiciones económicas continúan estables o mejoran
Productos • Custodios, supervisores y directores capacitados • Hogares de atención diurna modernizados y en operación • Materiales desarrollados • Sistema administrativo en buen funcionamiento • Sistema de información para la administración (MIS) en buen funcionamiento	• Custodios capacitados • Hogares modernizados y en operación • Materiales creados y distribuidos • MIS en funcionamiento	Datos del MIS sobre capacitados, hogares y materiales Evaluaciones de los capacitados Después de capacitación inicial y durante el curso de la capacitación continua	Las condiciones familiares permiten a las madres de los hogares de servicios cumplir sus acuerdos de prestación de servicios de cuidados.
Actividades • Seleccionar custodios y supervisores y ofrecer capacitación inicial • Modernizar los hogares • Desarrollar materiales • Desarrollar sistema administrativo • Prestar servicio diurno de cuidados • Ofrecer capacitación y supervisión continuas. • Desarrollar un MIS	*Recursos* Presupuesto Tecnología Recursos humanos	Plan de acción, presupuestos y registros contables Estudios que contengan el modelo escogido y trabajo del plan de capacitación Evaluaciones para determinar qué actividades se realizaron bien	

Fuente: Banco Interamericano de Desarrollo (http://www.iadb.org/sds/soc/eccd/6example.html#ex1).

08/11/2010 03:27:46 p.m.

La teoría del cambio debería estar en capacidad de responder las preguntas siguientes:

- ¿Se describe al programa en forma precisa con el modelo?
- ¿Están bien definidos todos los elementos?
- ¿Existe algún vacío en la cadena lógica de los eventos?
- ¿Son los elementos suficientes y necesarios?
- ¿Son factibles y consistentes las relaciones?
- ¿Es realista asumir que el programa produzca el logro de las metas declaradas de manera significativa?

La Kellog Foundation (2004) desarrolló una lista de verificación para evaluar la calidad de un modelo lógico. La siguiente lista de criterios es una adaptación de esa lista:

- Lista de las principales actividades requeridas para implementar el programa.
- Las actividades están conectadas claramente con la teoría programática especificada.
- Lista de los principales recursos necesarios para implementar el programa.
- Los recursos son los adecuados para el tipo de programa.
- Todas las actividades tienen recursos suficientes y apropiados.

Resumen

Los evaluadores deben resistirse al afán de apresurarse a una evaluación sin un plan. Un análisis del estado inicial constituye una parte valiosa de la planeación que ayuda a los evaluadores a obtener un cuadro más amplio del proyecto, el programa o la política. El análisis del estado inicial puede dar respuesta a preguntas importantes sobre la evaluación, incluyendo la oportunidad, el tiempo faltante para la conclusión, el personal participante, los recursos, el diseño, la teoría y lógica del programa y el conocimiento existente.

Una parte del análisis del estado inicial es identificar las personas o grupos que participan en la intervención. El análisis de las partes interesadas es una forma de identificar a las partes interesadas clave en la evaluación y determinar lo que conocen, su contribución y lo que esperan de la evaluación.

Otra parte del análisis del estado inicial es tener en cuenta el contexto de la intervención. Los evaluadores deben identificar y aprender de la investigación relacionada y de evaluaciones de intervenciones similares. Los evaluadores hacen distintas preguntas en diferentes etapas del ciclo de vida de una intervención, de modo que es importante identificar la etapa correspondiente.

La elaboración de una teoría del cambio subyacente a una intervención ayuda a los evaluadores y a las partes interesadas a visualizar la intervención, y determinar los supuestos subyacentes en el programa que deberán ser examinados en la evaluación. Aunque existen diferentes formas de describir gráficamente una teoría del cambio, todas deben basarse en los hallazgos de investigación y las evaluaciones relacionadas, describir el flujo lógico, identificar los eventos que pueden influir en los resultados y mostrar una cadena causal de eventos. Usualmente, las teorías del cambio describen insumos, actividades, productos y resultados, al igual que sus interrelaciones y el entorno en donde ocurren.

001El camino.indd 158 08/11/2010 03:27:47 p.m.

Actividades del capítulo 4

Actividad 4.1: aplicación de la teoría del cambio

Considérese un programa de microcréditos dirigido a promover nuevos medios de vida y mejorar el bienestar de las familias ayudando a las mujeres, en particular a las pobres del sector rural, a ingresar a la fuerza laboral y adquirir destrezas empresariales, para así aumentar el ingreso familiar. La meta a largo plazo es promover el desarrollo del sector privado e incrementar el crecimiento económico. El promedio de los préstamos es de US$225 con un monto máximo de US$500. Los préstamos se entregan como una suma global para capital de trabajo o inversión en una microempresa. Los plazos son entre 1 y 10 años, con un vencimiento promedio de 2 a 3 años, y se ofrece un período de gracia de un año. El componente asociado de formación de capacidad cubre la contabilidad y la administración financiera básicas. En el gráfico A se presenta un modelo simplificado de la lógica que describe el programa de microcréditos. En el gráfico B se presenta una descripción esquematizada más detallada de la teoría del cambio para el mismo programa. Los elementos encerrados en un círculo en el gráfico B muestran algunos factores importantes en el entorno que pueden influir en el logro de la meta del programa. Usted debe pensar en los supuestos subyacentes en el programa.

Enumere cinco supuestos de este programa:

1. _____
2. _____
3. _____
4. _____
5. _____

Gráfico 4.A *Teoría del cambio simple para un programa de micropréstamos*

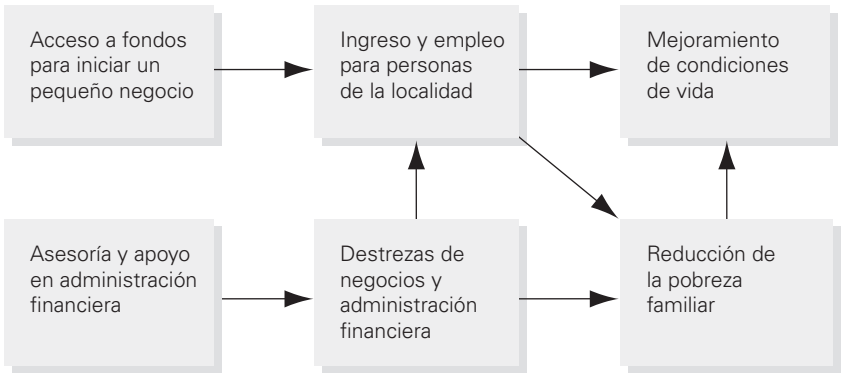
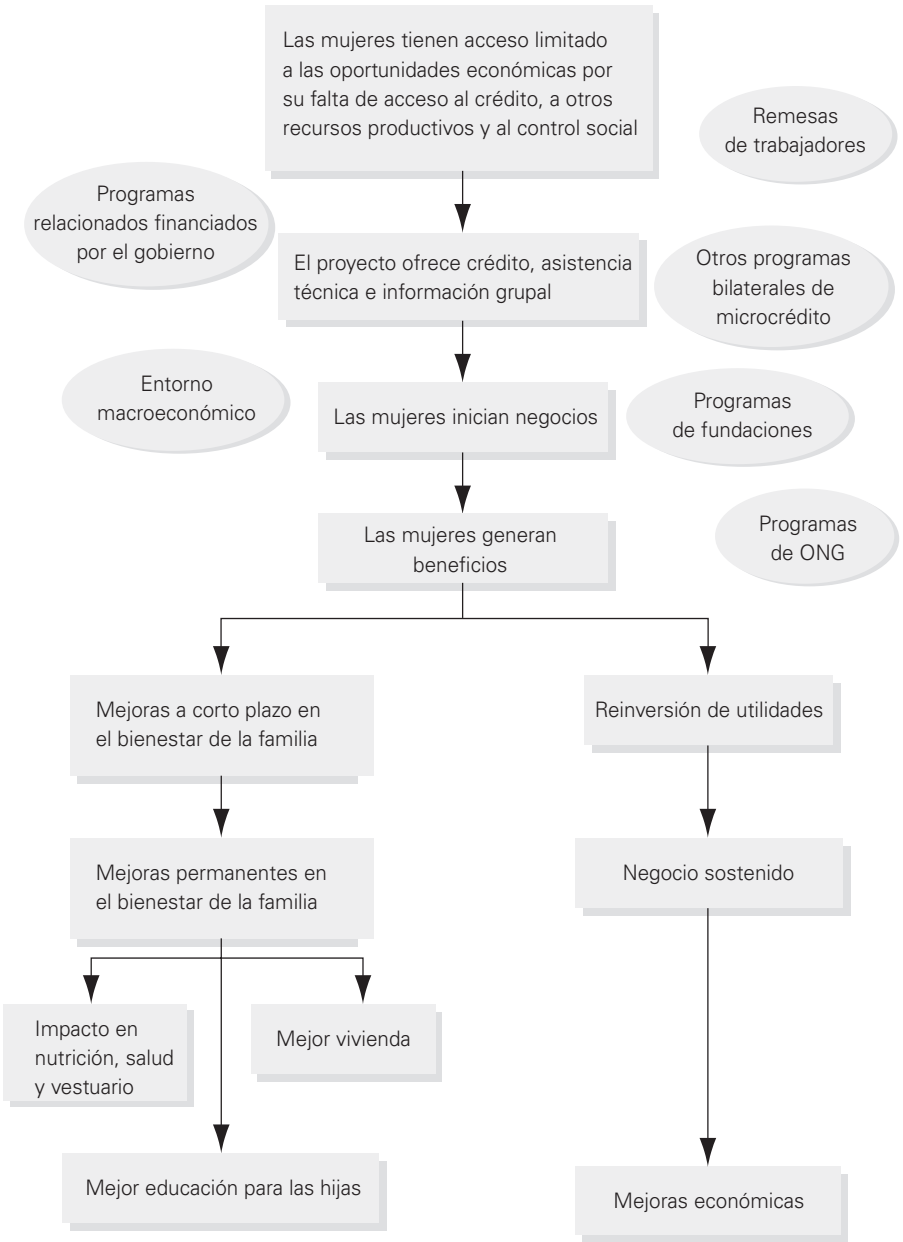


Gráfico 4.B *Teoría del cambio más complejo para un programa de micropréstamos*



Ejercicio de aplicación 4.2: analizar su programa

Considere un programa o proyecto en el que usted trabaja actualmente, o con el cual esté familiarizado pero no participando directamente en él. Desarrolle un gráfico que muestre la teoría del cambio para este programa o proyecto. Asegúrese de identificar los supuestos subyacentes, especialmente los relativos a los factores ambientales externos.

Referencias y lecturas adicionales

ActKnowledge and Aspen Institute. 2003. *Theory of Change.* Roundtable on Community Change. http://www.theoryofchange.org/.

Anderson, Andrea, A. 2004. "The Community Builder's Approach to Theory of Change: A Practical Guide to Theory Development". Presentación en la Aspen Institute Roundtable on Community Change, Nueva York. http://www.aspeninstitute.org/atf/cf/ percent7BDEB6F227-659B-4EC8-8F84-8DF23CA704F5 percent7D/rcccommbuildersapproach.pdf.

Banco Mundial. 1996. *World Bank Participation Sourcebook.* Washington, DC: Banco Mundial. http://www.worldbank.org/wbi/sourcebook/sbhome.htm.

———. 2004. *Monitoring and Evaluation: Some Tools, Methods and Approaches.* Operations Evaluation Department/Evaluation Capacity Development, Washington, DC. http://lnweb18.worldbank.org/oed/oeddoclib.nsf/b57456d58aba40e585256ad400736404/a5efbb5d776b67d285256b1e0079c9a3/$FILE/MandE_tools_methods_approaches.pdf.

Bell, P. 1997. "Using Argument Representations to Make Thinking Visible for Individuals and Groups". En *Proceedings of CSCL '97: The Second International Conference on Computer Support for Collaborative Learning,* eds. R. Hall, N. Miyake y N. Enyedy, 10-19. Toronto: University of Toronto Press.

Bruning, R. H., G. J. Schraw, M. M. Norby y R. R. Ronning. 2004. *Cognitive Psychology and Instruction.* 4ª ed. Upper Saddle River, NJ: Pearson Merrill Prentice-Hall.

Canadian Department of Foreign Affairs and International Trade. 2005. *Six Easy Steps to Managing for Results: A Guide for Managers.* Abril. Evaluation Division, Ottawa.

CIDA (Canadian International Development Agency). 2005. *Case Study #1: Jamaica Environmental Action Program ENACT.* Caribbean Division, Americas Branch. http://www.acdi-cida.gc.ca/CIDAWEB/acdicida.nsf/En/EMA-218131811-PHY#1.

Eggen, P. y D. Kauchak. 1999. *Educational Psychology: Windows on Classrooms.* 4ª ed. Upper Saddle River, NJ: Merrill: Prentice-Hall.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* Nueva York: Person Education.

Funnell, S. 1997. "Program Logic: An Adaptable Tool for Designing and Evaluating Programs". *Evaluation News and Comment* 6 (1): 5-7.

Gagne, R. M. y W. D. Rohwer Jr. 1969. "Instructional Psychology". *Annual Review of Psychology* 20: 381-418.

Gasper, D. 1997. "Logical Frameworks: A Critical Assessment Managerial Theory". Pluralistic Practice, documento de trabajo 264, Institute of Social Studies, La Haya.

———. 2000. "Evaluating the 'Logical Framework Approach' towards Learning-Oriented Development Evaluation". *Public Administration Development* 20 (1): 17-28.

Haarhuis, Carolien Klein. 2005. "Promoting Anticorruption Reforms: Evaluating the Implementation of a World Bank Anticorruption Program in Seven African Countries". http://igitur-archive.library.uu.nl/dissertations/2005-0714-200002/full.pdf.

Healy, A. F y D. S. McNamara. 1996. "Verbal Learning Memory: Does the Modal Model Still Work?" *Annual Review of Psychology* 47: 143-72.

Heubner, T. A. 2000. "Theory-Based Evaluation: Gaining a Shared Understanding Between School Staff and Evaluators". En *Program Theory in Evaluation: Challenges and Opportunities,* eds. Patricia J. Rogers, T. A. Hacsi, A. Petrosino y T. A. Huebner, 79-90. New Directions for Evaluation N° 87. San Francisco, CA: Jossey-Bass.

Kassinove, H. y M. Summers. 1968. "The Developmental Attention Test: A Preliminary Report on an Objective Test of Attention". *Journal of Clinical Psychology* 24 (1): 76-78.

Kellogg Foundation. 2004. *Logic Model Development Guide.* Battle Creek, MI. http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf.

Leeuw, Frans, Ger Gils y Cora Kreft. 1999. "Evaluating Anti-Corruption Initiatives: Underlying Logic and Mid-Term Impact of a World Bank Program". *Evaluation* 5 (2): 194-219.

Leeuw, Frans L. 1991. "Policy Theories, Knowledge Utilization y Evaluation". *Knowledge and Policy* 4: 73-92.

———. 2003. "Reconstructing Program Theories: Models Available and Problems to Be Solved". *American Journal of Evaluation* 24 (1): 5-20.

Mikkelsen, B. 2005. *Methods for Development Work and Research*: *A New Guide for Practitioners.* Thousand Oaks, CA: Sage Publications.

MSH (Management Sciences for Health) y UNICEF (Fondo de las Naciones Unidas para la Infancia). 1998. "Quality Guide: Stakeholder Analysis". En *Guide to Managing for Quality.* http://ERC.MSH.org/quality.

Newman, David Kent. 2007. "Theory: Write-up for Conceptual Framework". http://deekayen.net/theory-write-conceptual-framework.

Ormrod, J. E. 2006. *Essentials of Educational Psychology.* Upper Saddle River, NJ: Pearson Merrill Prentice-Hall.

Owen, J. M. y P. J. Rogers. 1999. *Program Evaluation: Forms and Approaches.* Thousand Oaks, CA: Sage Publications.

Pancer, S. Mark y Anne Westhues. 1989. "A Developmental Stage Approach to Program Planning and Evaluation". *Evaluation Review* 13 (1): 56-77.

Patton, M. Q. 2002. *Qualitative Research and Evaluation Methods.* 3ª ed. Thousand Oaks, CA.: Sage Publications.

Pawson, Ray. 2006. *Evidence-Based Policy: A Realistic Perspective.* Nueva Brunswick, NJ: Sage Publications.

Petrosino, Anthony, Robert A. Boruch, Cath Rounding, Steve McDonald y Lain Chalmers. 2003. "The Campbell Collaboration Social, Psychological, Educational, and Criminological Trials Register C20SPECTR". http://www.campbellcollaboration.org/papers/unpublished/petrosiN°pdf.

Porteous, Nancy L., B. J. Sheldrick y P. J. Stewart. 1997. *Program Evaluation Tool Kit: A Blueprint for Public Health Management.* Ottawa-Carleton Health Department, Ottawa. http://www.phac-aspc.gc.ca/php-psp/tookit.html.

———. 2002. "Introducing Program Teams to Logic Models: Facilitating the Learning Process". *Canadian Journal of Program Evaluation* 17 (3): 113-41.

Prensky, Marc. 2001. "Digital Natives, Digital Immigrants". *On the Horizon* 9 (5). http://www.marcprensky.com/writing/Prensky percent20-percent20Digitalpercent20Natives,percent20Digital percent20Immigrants percent20-percent20Part1.pdf.

Ritzer, George. 1993. *The McDonaldization of Society.* Rev. ed. Thousand Oaks, CA: Pine Forge Press.

Rogers, Patricia J., T. A. Hacsi, A. Petrosino y T. A. Huebner, eds. 2000. *Program Theory in Evaluation: Challenges and Opportunities.* New Directions in Evaluation N° 87. San Francisco: Jossey-Bass Publishers.

Rossi, P., H. Freeman y M. Lipsey. 1999. *Evaluation: A Systematic Approach.* Thousand Oaks, CA: Sage Publications.

Scott, M. 2003. "The Benefits and Consequences of Police Crackdowns". *Problem-Oriented Guides for Police, Response Guide 1.* U.S. Department of Justice, Office of Community-Oriented Policing Services, Washington, DC.

Scriven, Michael. 2007. *Key Evaluation Checklist*. http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf.

Shadish, W. R. Jr., T. D. Cook y L. C. Leviton. 1991. *Foundations of Program Evaluation.* Thousand Oaks, CA: Sage Publications.

Sherman, L. W., D. Farrington, B. C. Welsh y D. L. MacKenzie, eds. 2002. *Evidence-Based Crime Prevention.* Londres: Routledge.

Stufflebeam, Daniel L. 2001. *Evaluation Models*. New Directions for Evaluation N° 89. San Francisco, CA: Jossey-Bass.

Stufflebeam, D. L., G. F Madaus y T. Kellaghan, eds. 2000. *Evaluation Models.*2ª ed. Boston: Kluwer Academic Publishers.

Suthers, D. D. y A. Weiner. 1995. *Belvédère.* http://lilt.ics.hawaii.edu/belvedere/index.html.

Suthers, D. D., E. E. Toth y A. Weiner 1997. "An Integrated Approach to Implementing Collaborative Inquiry in the Classroom". En *Proceedings of CSCL '97: The Second International Conference on Computer Support for Collaborative Learning,* eds. R. Hall, N. Miyake y N. Enyedy, 272-79. Toronto: University of Toronto Press.

Swedberg, Richard. 2003. *Principles of Economic Sociology.* Princeton, NJ: Princeton University Press.

Taylor-Powell, Ellen. 2005. *Logic Models: A Framework for Program Planning and Evaluation.* University of Wisconsin Extension, Program Development and Evaluation. http://www.uwex.edu/ces/pdande/evaluation/pdf/nutritionconf05.pdf.

U.S. GAO (General Accounting Office). 1991. *Designing Evaluations.* Washington, DC. http://www.gao.gov/special.pubs/10_1_4.pdf.

Weiss, Carol H. 1997. *Evaluation: Methods for Studying Programs and Policies.* Upper Saddle River, NJ: Prentice-Hall.

Worthen, Blaine. R., James R. Sanders y Jody L. Fitzpatrick. 1997. *Program Evaluation.* Nueva York: Longman.

Sitios web

Campbell Collaboration. http://www.campbellcollaboration.org/.

CIID (Centro Internacional de Investigaciones para el Desarrollo). 2004. *Evaluation Planning in Program Initiatives* Ottawa, Ontario, Canada. http://web.idrc.ca/uploads/user-S/108549984812guideline-web.pdf.

Community Toolbox. *A Framework for Program Evaluation: A Gateway to Tools.* http://ctb.lsi.ukans.edu/tools/EN/sub_section_main_1338.htm.

Evaluation Center, Western Michigan University. *The Checklist Project.* http://www.wmich.edu/evalctr/checklists/checklistmenu.htm#mgt.

Suthers, D. y A. Weiner. 1995.*Groupware for Developing Critical Discussion Skills.* http://www-cscl95.indiana.edu/cscl95/suthers.html.

University of Wisconsin-Extension. *Logic Model.* http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html.

———. *Examples of Logic Models.* http://www.uwex.edu/ces/pdande/evaluation/evallogicmodelexamples.html.

Teorías del cambio

Audience Dialogue. 2007a. "Enhancing Program Performance with Logic Models". http://www.uwex.edu/ces/lmcourse/, http://www.wkkf.org/pubs/tools/evaluation/pub3669.pdf.

———. 2007b. "Program Logic: An Introduction". http://www.audiencedialogue.org/proglog.htm.

AusAid. 2005. "Section 3.3: The Logical Framework Approach and Section 2.2: Using the Results Framework Approach". http://www.ausaid.gov.au/ausguide/default.cfm.

Banco Interamericano de Desarrollo. http://www.iadb.org/sds/soc/eccd/bexample.html#ex1.

BOND. 2001. *Guidance Notes Series I: Beginner's Guide to Logical Framework Analysis.* http://www.ngosupport.net/graphics/NGO/documents/english/273_BOND_Series_1.doc.

Davies, Rick. 2003. "Network Perspective in the Evaluation of Development Interventions: More than a Metaphor". Documento presentado en la EDAIS Conference "New Directions in Impact Assessment for Development: Methods and Practice", Noviembre 24-25. http://www.enterprise-impact.org.uk/conference/Abstracts/Davies.shtml and http://www.mande.co.uk/docs/nape.doc.

Department for International Development. 2002. *Tools for Development: A Handbook for Those Engaged in Development Activity Performance and Effectiveness.* http://www.dfid.gov.uk/pubs/files/toolsfordevelopment.pdf.

Den Heyer, Molly. 2001a. *A Bibliography for Program Logic Models/Logframe Analysis.* Evaluation Unit, International Development Research Centre. http://www.idrc.ca/uploads/user-S/10553606170logframebib2.pdf and http://www.mande.co.uk/docs/Phillips.ppt.

———. 2001b. "The Temporal Logic Model™: A Concept Paper". http://www.idrc.ca/uploads/user-S/10553603900tlmconceptpaper.pdf.

International Fund for Agricultural Development. 2003. *Annotated Example of a Project Logframe Matrix.* http://www.ifad.org/evaluation/guide/3/3.htm and http://www.ifad.org/evaluation/guide/annexb/index.htm.

JISC infoNet. 2001. *Engendering the Logical Framework.* http://www.jiscinfonet.ac.uk/InfoKits/project-management/InfoKits/infokit-related-files/logicalframework-nformation.

———. 2004. *Logical Framework (LogFRAME) Methodology.* http://www.jiscinfonet.ac.uk/InfoKits/project-management/InfoKits/infokit-related-files/logicalframework-information.

Kellogg Foundation. 2004. *Evaluation Logic Model Development Guide: Using Logia Models to Bring Together Planning, Evaluation y Action.* http://www.wkkf.org./Pubs/Tools/Evaluation/Pub3669.pdf.

Knowledge and Research Programme on Disability and Healthcare Technology. 2004. *Constructing a Logical Framework.* http://www.kar-dht.org/logframe.html.

McCawley, Paul F. 1997. *The Logic Model for Program Planning and Evaluation.* University of Idaho Extension. http://www.uidaho.edu/extension/LogicModel.pdf.

NORAD (Norwegian Agency for Development Cooperation). 1999. *The Logical Framework Approach: Handbook for Objectives-Oriented Planning.* 4ª ed. http://www.norad.no/default.asp?V_ITEM_ID=1069.

PPM&E Resource. 2005. *Logical Framework Approach.* portals.wdi.wur.nl/ppme/index.php?Logical_Framework_Approach.

Rugh, Jim. "The Rosetta Stone of Logical Frameworks". http://www.mande.co.uk/docs/Rosettastone.doc.

SIDA (Agencia Sueca de Cooperación para el Desarrollo Internacional). 2004. *The Logical Framework Approach: A Summary of the Theory behind the LFA Method.* http://www.sida.se/shared/jsp/download.jsp?f=SIDA1489en_web.pdf&a=2379.

Usable Knowledge, USA. 2006. Online Logic Model Training: An Audiovisual Presentation. http://www.usablellc.net/Logic percent20Modelpercent20percent28Online percent29/Presentation_Files/index.html.

Weiss, Carol. 1972. *Evaluation Research: Methods for Assessing Program Effectiveness.* Englewood Cliffs, New Jersey: Prentice Hall.

Wikipedia. 2006. "Logical Framework Approach". http://en.wikipedia.org/wiki/Logical_framework_approach.

Críticas del marco lógico

Bakewell, Oliver y Anne Garbutt. 2006. *The Use and Abuse of the Logical Framework Approach A Review of International Development NGOs' Experiences.*

Swedish International Development Agency. http://www.sida.se/shared/jsp/download.jsp?f=LFA-review.pdf&a=21025.

MISEREOR. 2005. *Methodological Critique and Indicator Systems.* http://www.misereor.org/index.php?id=4495.

Versiones alternas del marco lógico

Broughton, Bernard. 2005. *A Modified Logframe for Use in Humanitarian Emergencies.* http://www.mande.co.uk/docs/EmergencyLogframeBroughton.doc.

Shaikh, Partha Hefaz. 2005. *Intertwining Participation, Rights-Based Approach and Log Frame: A Way Forward in Monitoring and Evaluation for Rights-Based Work.* http://www.mande.co.uk/docs/PIFA_Article_PDF.pdf. SIDA Civil Society Centre. 2006. *Logical Framework Approach, with an Appreciative Approach.* http://www.sida.se/shared/jsp/download.jsp?f=SIDA28355en_LFA_web.pdf&a=23355.

001El camino.indd 166 08/11/2010 03:27:47 p.m.

CAPÍTULO 5

Consideración sobre el enfoque de la evaluación

El trabajo para el desarrollo ha pasado de concentrarse en los proyectos a hacerlo en los programas y políticas, haciendo énfasis en la sostenibilidad. Se ha utilizado una amplia variedad de enfoques en el diseño y realización de evaluaciones para atender estas materias más amplias y complejas. En este capítulo se tratan algunos de estos enfoques.

Enfoques generales en la evaluación

Desde los años noventa, la comunidad internacional ha venido pasando lentamente del enfoque en los proyectos al enfoque en los programas y las asociaciones, en la ayuda para el desarrollo (*ver* cuadro 1.3). Los enfoques de asociación incluyen un mayor número de partes interesadas, que participan en un mayor grado en operaciones más complejas, como las de enfoques de financiación en todo un sector. En consecuencia, las evaluaciones de programas grandes ahora son más participativas y se realizan conjuntamente, si bien las evaluaciones conjuntas siguen siendo la norma. Las convocatorias renovadas por una mayor financiación no ligada a los gobiernos –por ejemplo, las expresadas en Agra en 2008– implican mayores retos para la evaluación en el futuro.

001El camino.indd 167 08/11/2010 03:27:47 p.m.

Se han desarrollado varios enfoques para satisfacer la naturaleza cambiante de la evaluación para el desarrollo. La elección del enfoque de evaluación depende parcialmente del contexto; los enfoques no necesariamente son excluyentes y las evaluaciones pueden combinar elementos de dos o más enfoques. Los enfoques incluyen los siguientes:

- Evaluación prospectiva.
- Evaluación de la evaluabilidad.
- Evaluación basada en metas.
- Evaluación sin tener en cuenta las metas.
- Evaluación en sitios múltiples.
- Evaluación de conjunto.
- Evaluación social.
- Evaluación ambiental y social.
- Evaluación participativa.
- Cartografía de los efectos directos.
- Evaluación rápida.
- Síntesis de evaluación y metaevaluación.
- Otros enfoques.

Cualquiera que sea el enfoque seleccionado, deben darse los mismos pasos de planeación: definir los temas de evaluación, identificar las medidas, recopilar y analizar los datos e informar y utilizar los hallazgos.

Evaluación prospectiva

Evaluación prospectiva: De los probables efectos directos de un proyecto, un programa o una política propuesta.

Una evaluación prospectiva se realiza *ex ante*; esto es, el programa propuesto se revisa antes de empezar, para intentar analizar su probabilidad de éxito, predecir su costo y analizar propuestas y proyecciones alternativas. Las evaluaciones prospectivas han sido realizadas por organizaciones de evaluación dentro de la rama legislativa del poder. Un ejemplo de éstas es la U.S. Government Accountability Office (GAO) –su nuevo nombre desde 2004– que reporta al Congreso de los EUA. Los evaluadores de la GAO asisten en ocasiones a quienes toman las decisiones en el gobierno, suministrando información analítica sobre temas y opciones sobre programas potenciales (U.S. GAO, 1990).

Con mucha frecuencia se solicita información a la GAO sobre el probable éxito de los nuevos programas propuestos. La GAO revisa la información sobre propuestas alternativas y analiza los efectos directos de programas similares que pueda haber en curso o ya terminados. En el cuadro 5.1 se identifican cuatro clases de preguntas de enfoque al futuro que se solicita investigar a la GAO.

La mayoría de las evaluaciones prospectivas implican las actividades siguientes:

- Un análisis contextual del programa o política propuesta.

Cuadro 5.1 *Tipos de preguntas de enfoque al futuro y formas de tratarlas*

Propósito de la pregunta	*Criticar otros análisis*	*Efectuar un análisis propio*
Anticipar el futuro	¿Qué tan bien ha proyectado la administración las necesidades, los costos y consecuencias futuras?	¿Cuáles son las necesidades, costos y consecuencias futuras?
Mejorar las acciones futuras	¿Cuál es el éxito probable de una propuesta de la administración o del Congreso?	¿Qué curso de acción tiene mayor potencial de éxito?

Fuente: U.S. GAO, 1990.

- Una revisión de estudios de evaluación sobre programas o políticas similares y síntesis de los hallazgos y enseñanzas del pasado.
- Una predicción del éxito o el fracaso probable, dado un contexto futuro no muy distinto del pasado y sugerencias sobre el fortalecimiento del programa o política propuesta, si quienes toman las decisiones deciden seguir adelante (GAO, 1990).

(En la lista de sitios web, al final de este capítulo, se presentan recursos y ejemplos de los distintos tipos de evaluaciones).

Evaluación de la evaluabilidad

Evaluación de la evaluabilidad: Breve estudio preliminar que se efectúa para determinar si sería útil y factible una evaluación.

Una evaluación de la evaluabilidad es un breve estudio preliminar que se efectúa para determinar si sería útil y factible una evaluación. Esta clase de estudio preliminar ayuda a clarificar las metas y objetivos del programa o el proyecto, identificar los recursos de datos disponibles, precisar los vacíos, identificar datos que es necesario desarrollar, e identificar a las partes interesadas clave y clarificar sus necesidades de información. Puede también redefinir el propósito de la evaluación y los métodos de realizarla. Al considerar la intervención como si estuviera implementada en el terreno y las implicaciones para la oportunidad y el diseño de la evaluación, una evaluación de la evaluabilidad puede ahorrar tiempo y contribuir a evitar errores costosos.

Wholey y sus colegas desarrollaron el enfoque de la evaluación de la evaluabilidad, a principios de los años setenta, por su convicción de que muchas evaluaciones fracasaban debido a discrepancias entre la "retórica y la realidad" (Nay y Kay, 1982, p. 225). Ellos vieron la evaluación de la evaluabilidad como un medio de facilitar la comunicación entre los evaluadores y las partes interesadas, y propusieron utilizarla como medio para determinar si un programa era "evaluable" y para enfocar la evaluación (Fitzpatrick, Sanders y Worthen, 2004). La evaluación de la evaluabilidad se desarrolló originalmente como precursora de la evaluación sumativa; su papel se ha expandido desde entonces para incluir la clarificación de propósitos de un estudio formativo, o para funcionar como herramienta de planificación (Smith, 1989).

En una evaluación de la evaluabilidad, los evaluadores realizan un trabajo preliminar para determinar si puede efectuarse una evaluación. Si, por ejemplo, se propone una evaluación basada en metas (o en objetivos), puede ser problemática la falta de objetivos claros del programa o de acuerdo entre las partes interesadas. A veces, no hay medidas disponibles o son inaccesibles. En este caso deben desarrollarse antes de poder evaluar la intervención.

Así, la evaluación de la evaluabilidad se centra en la factibilidad de realizar una evaluación. Si no es factible diseñar una evaluación a partir de la información que está disponible, o la intervención carece de una teoría del cambio coherente, será necesario hacer más trabajo preliminar. Una evaluación de la evaluabilidad ayuda a clarificar la intervención y los efectos directos que se desea obtener de ella, y con mucha frecuencia genera un diálogo sobre las metas y los objetivos, los productos, la población objetivo y los efectos directos deseados de una intervención, de tal manera que se llegue a un acuerdo sobre lo que debe lograrse.

Las evaluaciones de evaluabilidad son hechas a menudo por grupos de partes interesadas, como ejecutores, administradores y evaluadores. Para realizar una evaluación de evaluabilidad, el equipo:

- Revisa materiales en los que se define y describe la intervención.
- Identifica modificaciones a la intervención.
- Entrevista a los administradores y al personal sobre su percepción de las metas y objetivos de la intervención.
- Entrevista a las partes interesadas sobre su percepción de las metas y los objetivos de la intervención.
- Desarrolla o refina un modelo de la teoría del cambio.
- Identifica las fuentes de datos.
- Identifica personas y organizaciones que puedan poner en práctica posibles recomendaciones de la evaluación.

Uno de los beneficios potenciales de la evaluación de la evaluabilidad es que puede llevar a una evaluación más realista y apropiada. Smith (1989) y Wholey (1987) señalan además que puede mejorar:

- La capacidad de distinguir entre el fracaso en la intervención y el fracaso en la evaluación.
- La estimación de los efectos directos a largo plazo.
- La inversión de las partes interesadas en la intervención.
- El desempeño de la intervención.
- El desarrollo de la intervención y las destrezas del personal de evaluación.
- La visibilidad y la rendición de cuentas por la intervención.
- La comprensión administrativa de la intervención.
- Las elecciones de políticas.
- El apoyo continuo a la intervención.

Evaluación basada en metas

Una evaluación basada en metas mide el grado al cual un programa o una intervención logra objetivos claros y específicos (Patton, 1990). El foco de la evaluación está en los efectos directos declarados (las metas o los objetivos) del proyecto, el programa o la política. Esta es la evaluación usual familiar para la mayor parte de las personas, y es la base de la mayoría de los sistemas de evaluación de proyectos de las organizaciones para el desarrollo.

Una crítica del enfoque basado en los objetivos es que se concentra en los aspectos económicos y técnicos, en lugar de los aspectos sociales y humanos (Hirschheim y Smithson, 1988). Una segunda crítica es la de que tales evaluaciones se centran sólo en los objetivos declarados, y no se evalúa el logro de otros objetivos importantes, que pueden ser implícitos y no explícitos, o pueden haberse tratado en las reuniones de la junta o de la división, pero no se incluyeron en los objetivos declarados del proyecto. Una tercera crítica relacionada es que no buscan efectos no previstos (positivos o negativos).

Evaluación basada en metas: Mide el grado al cual un programa o una intervención logra objetivos claros y específicos.

Estas son omisiones serias. Una evaluación de una nueva planta de tratamiento de agua, por ejemplo, puede centrarse en los objetivos declarados en el proyecto de construir, operar y mantener una instalación de tratamiento de agua para producir agua limpia en un determinado volumen por día, para satisfacer las necesidades de determinada cantidad de familias o comunidades. A menudo, se pasan por alto los efectos en las personas que viven en el terreno que deben trasladarse a otro sitio antes de iniciarse la construcción. El fracaso en articular una meta o un objetivo para el traslado de esas personas, de modo que obtengan un medio de vida mejor y sostenible, representa una omisión seria. La evaluación agrava el problema si no hace cuestionamientos sobre el traslado porque no fue un objetivo formal y explícito del proyecto.

La evaluación debería buscar efectos positivos o negativos no previstos de la intervención. Aunque no se haya articulado un objetivo específico, sería útil que la evaluación documentara la menor incidencia de enfermedades transmitidas por el agua en las comunidades que ahora la reciben potable.

La evaluación basada en resultados, el método que se defiende en este texto, busca resultados, estén o no articulados como metas u objetivos. Las evaluaciones basadas en objetivos pueden fortalecerse abriéndolas a los resultados imprevistos, positivos o negativos.

Evaluación sin tener en cuenta las metas

La evaluación sin tener en cuenta las metas se desarrolló como una reacción a la evaluación basada en objetivos. Scriven (1972b), quien propuso inicialmente la evaluación sin tener en cuenta las metas, ha sido un mayor defensor de este enfoque.

Evaluación sin tener en cuenta las metas: Evaluación en la cual los evaluadores deliberadamente intentan evitar toda retórica relativa a los objetivos del programa, basando la evaluación solamente en el grado al cual satisface las necesidades de los participantes.

En la evaluación sin tener en cuenta las metas el evaluador intenta deliberadamente evitar toda retórica relativa a los objetivos del programa. Los evaluadores no discuten

los objetivos con el personal, ni leen folletos o propuestas del programa. Evalúan sólo los efectos directos observables y los documentables en relación con las necesidades de los participantes (Patton, 2002). Como anota Scriven (1972b):

> En resumen, me pareció que la consideración y la evaluación de los objetivos era un paso innecesario, pero posiblemente también contaminante… Cuanto menos oiga el evaluador externo sobre los objetivos del proyecto, menos visión de túnel desarrollará y prestará más atención a la búsqueda de efectos reales (en lugar de revisar efectos alegados).

En las evaluaciones sin tener en cuenta las metas se recopilan datos sobre los efectos y la eficacia de los programas, sin estar restringidas por un enfoque estrecho en metas y objetivos. En dichas evaluaciones se capturan las experiencias reales de los participantes en el programa en sus propios términos. Requieren que los evaluadores suspendan el juicio sobre lo que un programa intenta hacer y centrarse en cambio en averiguar qué ocurre en la realidad. Por estas razones es especialmente compatible con la investigación cualitativa, si bien puede emplear métodos cuantitativos y también cualitativos.

Scriven (1997) propone hacer evaluaciones separadas, sin tener en cuenta las metas y basadas en objetivos, para maximizar las fortalezas y minimizar las debilidades de cada enfoque.

Wholey, Harty y Newcomer (1994) describen las siguientes características de la evaluación libre de metas:

- El evaluador evita conocer los objetivos del programa.
- No se permite que objetivos predeterminados estrechen el foco del estudio de evaluación.
- El evaluador tiene un contacto mínimo con el administrador del programa y el personal.
- El foco se sitúa en los efectos directos reales del programa y no en los deseados.

Por lo general, en el contexto del desarrollo es difícil evitar conocer las metas y los objetivos de una intervención. Puede no ser factible o deseable que el evaluador tenga un contacto mínimo con el personal del programa. No obstante, al adoptar un enfoque basado en resultados, el evaluador puede utilizar muchos de los elementos de una evaluación libre de metas.

Evaluación en sitios múltiples

En una intervención en gran escala es necesario con mucha frecuencia observar las intervenciones ejecutadas en varias ubicaciones. Esto se conoce como evaluaciones en sitios múltiples. La evaluación puede haberse implementado en la misma forma en todas las ubicaciones o de manera diferente en algunas de ellas. Una evaluación en sitios múltiples suministra información sobre la experiencia general de la intervención, e igualmente un conocimiento más profundo de las variaciones que ocurrieron. Responde preguntas como estas:

Evaluación en sitios múltiples: Las intervenciones ejecutadas en varias ubicaciones.

- ¿Qué características de la implementación de la intervención son comunes a todas las ubicaciones?
- ¿Qué características varían y por qué?
- ¿Se basan las diferencias en los efectos directos en esas variaciones?

Aquí un asunto clave es obtener información en profundidad. Para conseguirlo, los evaluadores utilizan a menudo casos de estudio para la evaluación en sitios múltiples. Los sitios se seleccionan generalmente para el estudio, porque representan ciertas características (por ejemplo, tamaño, etnia, condición socioeconómica) que pueden resultar en diferencias sistémicas en la implementación de la intervención y los efectos directos. Desde luego, puede ser difícil determinar si las variaciones en la intervención fueron las que causaron las diferencias. En algunos casos, las intervenciones muestran impactos debido a diferencias únicas en un entorno, tales como un fuerte liderazgo o una ciudadanía activa. En otros casos, los cambios pueden explicarse por diferencias sistemáticas, por ejemplo diferencias regionales. Éstas pueden tener implicaciones para las réplicas de la intervención.

Una evaluación en sitios múltiples debe captar el clima en el que operan las intervenciones, lo mismo que cualquier diferencia sistemática cultural, geográfica, económica, de tamaño o de otra clase, que pueda afectar la variación en las experiencias y efectos directos. La participación de las partes interesadas es importante, pues éstos pueden ayudarle al evaluador a entender mejor la situación local.

Una evaluación en sitios múltiples es usualmente más fuerte que una evaluación de una sola intervención en una sola ubicación. Puede resumir en forma más verosímil una población mayor, porque incluye una muestra más grande y un conjunto más diverso de situaciones de intervención. Puede efectuar análisis "dentro de" y también "entre". Los hallazgos generales y también los consistentes a través de intervenciones, ofrecen una evidencia más fuerte de la eficacia de una intervención que la obtenida en un solo sitio.

Es probable que la comparación de una intervención en varios sitios aporte varias enseñanzas y estrategias para tratar con varias situaciones. Las buenas prácticas pueden también surgir a partir de una evaluación en sitios múltiples. Sin embargo, es importante tener presente que los sitios seleccionados con base en el juicio del evaluador, aun cuando se seleccionen por su representación de determinadas características, no son estadísticamente representativos de la población y no necesariamente revelan buenas prácticas.

Al efectuar evaluaciones de sitios múltiples, los evaluadores deben asegurarse de que la recopilación de datos sea lo más estandarizada posible. Los mismos datos deben recopilarse en lo posible en la misma forma para que las comparaciones puedan ser significativas. La recopilación requiere personal bien capacitado, disponibilidad de los mismos datos en cada sitio e información previa suficiente para diseñar los instrumentos de recopilación de los datos.

Cada ubicación es diferente y, aunque algunos indicadores pueden ser comparables, cada sitio puede haber tenido un enfoque ligeramente distinto. Los contextos político,

social, económico e histórico conforman la implementación del proyecto y, por consiguiente, la evaluación (Johnston y Stout, 1999).

Evaluación de conjunto

Una evaluación de conjunto es similar a una evaluación de sitios múltiples, pero en este caso la intención es diferente. Por lo general, en ella se examinan grupos de intervenciones similares o relacionadas que juntas representan un conjunto de intervenciones. En ella se tiene en cuenta este "emporio" de intervenciones en uno o más entornos. A semejanza de una evaluación en sitios múltiples, una evaluación de conjunto se concentra en intervenciones que comparten una misión, una estrategia y una población objetivo comunes. A diferencia de una evaluación en sitios múltiples, el propósito de una evaluación de conjunto no es determinar si una intervención funciona o asegurar la rendición de cuentas. No evalúa el éxito o fracaso de intervenciones individuales ni identifica las intervenciones que deben terminarse. Su propósito es aprender sobre lo ocurrido en los grupos y verificar temas comunes y enseñanzas aprendidas. La información sólo se presenta en forma agregada, de manera que no se identifica ningún proyecto individual. Al igual que en las evaluaciones en sitios múltiples, la participación de las partes interesadas es básica.

Evaluación de conjunto: Evaluación en la que se examinan grupos de intervenciones similares o relacionadas.

Las evaluaciones grupales difieren de las evaluaciones en sitios múltiples en que no se preocupan por la posibilidad de generalizar o de replicar las intervenciones. La variación se considera como positiva porque los proyectos individuales se ajustan a su contexto particular. Una evaluación de conjunto se concentra más en el aprendizaje que en sacar conclusiones generales sobre la calidad o el valor del programa.

Aunque no existe una metodología específica para las evaluaciones grupales, en éstas a menudo se utilizan enfoques cualitativos para complementar los datos cuantitativos. Las evaluaciones grupales se pueden ver como casos de estudio múltiples, con información compartida entre los casos a través de conferencias por redes como característica significativa de este enfoque.

Una desventaja de las evaluaciones grupales es que no presentan efectos directos para los sitios individuales ni tienen en cuenta la variación planeada o imprevista. Los datos sólo muestran información agregada.

Evaluación social

La evaluación social se ha convertido en parte importante de muchas evaluaciones. En una evaluación social se tienen en cuenta las estructuras, los procesos y cambios sociales dentro de un grupo o una comunidad, y se pueden también tener en cuenta las tendencias que puedan afectar al grupo.

Una evaluación social es el principal instrumento utilizado para garantizar que se tengan en cuenta los impactos sociales de los proyectos, programas y políticas para

Evaluación social: Evaluación en la que se examinan las estructuras, los procesos y cambios sociales dentro de un grupo o una comunidad.

el desarrollo. Es utilizada para comprender los asuntos y riesgos sociales clave, y para determinar el impacto social de una intervención en las diferentes partes interesadas. En particular, las evaluaciones sociales están dirigidas a establecer si es probable que un proyecto cause impacto social adverso (por ejemplo, desplazar residentes para construir una planta de energía). Pueden adoptarse estrategias para mitigar el impacto adverso si éste se conoce y es reconocido, y con las estrategias de mitigación se puede entonces controlar y evaluar como parte de la evaluación.

En el *World Bank Participation Sourcebook* (1996) se identifican los siguientes fines de la evaluación social:

- Identificar las principales partes interesadas y establecer un marco apropiado para su participación en la selección, el diseño y la ejecución del proyecto.
- Asegurar que los objetivos del proyecto y los incentivos para el cambio sean aceptables para la variedad de personas que se espera beneficiar, y que se reflejen las diferencias de género y otras diferencias sociales en el diseño del proyecto.
- Evaluar el impacto social de los proyectos de inversión, donde se identifiquen impactos adversos, determinar cómo pueden superarse o al menos mitigarse en forma sustancial.
- Desarrollar capacidad en el nivel apropiado para habilitar la participación, resolver los conflictos, permitir la prestación del (los) servicio(s), y adoptar medidas de mitigación según sea necesario.

En el *Participation Sourcebook* se identifican las siguientes preguntas comunes que se hacen durante la evaluación social:

- ¿Quiénes son las partes interesadas? ¿Son consistentes los objetivos del proyecto con sus necesidades, intereses y capacidades?
- ¿Qué factores sociales y culturales afectan la capacidad de las partes interesadas de participar o de recibir beneficio de las operaciones propuestas?
- ¿Cuál es el impacto del proyecto o el programa en las distintas partes interesadas, en particular las mujeres y los grupos vulnerables? ¿Cuáles son los riesgos sociales (falta de compromiso o capacidad e incompatibilidad con las condiciones existentes) que pueden afectar el éxito del proyecto o el programa?
- ¿Qué acuerdos institucionales son necesarios para la participación y la ejecución del proyecto? ¿Existen planes adecuados para formar la capacidad requerida para cada uno de ellos?

Las herramientas y los enfoques de evaluación incluyen los siguientes:

- Análisis de las partes interesadas.
- Análisis de género.
- Evaluación de participación rural.
- Observación, entrevistas y grupos de discusión.

- Cartografía, análisis de tareas y clasificación de la riqueza.
- Talleres centrados en la planeación de proyectos orientada a objetivos.

Los ejemplos de indicadores básicos para el seguimiento del impacto social incluyen los siguientes:

- Tasas de participación de grupos sociales en pruebas voluntarias
- Tasas de participación de grupos sociales en actividades de consejería
- Informes de mayor uso de preservativos
- Porcentaje de integrantes de la comunidad que participa en la atención a personas con VIH/Sida y sus familias.

En el recuadro 5.1 se ofrece un ejemplo de la incorporación de la evaluación social en un proyecto.

Evaluación ambiental y social

Con mayor frecuencia, las organizaciones para el desarrollo reconocen la necesidad de programas y proyectos para abordar los problemas ambientales y sociales, y para evaluar el logro de los objetivos relacionados con ellos. La mayor parte de las organizaciones para el desarrollo se adhieren a un conjunto de estándares centrales ambientales y sociales, e identifican impactos potenciales ambientales y sociales como parte del proceso de aprobación de un programa o un proyecto. De aprobarse el programa o el proyecto, estos impactos potenciales se controlan durante la ejecución y se evalúan a su terminación. Las evaluaciones ambientales y sociales se consideran inseparables.

Las organizaciones para el desarrollo reconocen ya el papel que debe tener la población local en el diseño, la implementación y el seguimiento de las intervenciones que tienen implicaciones para los recursos ambientales y naturales. La población local y otras partes interesadas son socios en la conservación y la administración de los recursos naturales.

La evaluación ambiental y social puede ser el único propósito de una evaluación, o puede ser un componente de la evaluación. Las evaluaciones ambientales y sociales deben realizarse no sólo en proyectos con efectos ambientales obvios (plantas de pulpa y papel, tuberías petroleras), sino también en intervenciones como la construcción de un nuevo colegio o la financiación de una línea de crédito.

Si una organización carece de estándares ambientales y sociales, los evaluadores pueden usar los estándares del país en donde funciona la organización, los Equator Principles, la ISO 14031 y el libro *Sustainable Development Strategies: A Resource Book* (Dalal-Clayton y Ba, 2002).

Los principios ecuatoriales

Son un conjunto de principios que ayudan a las entidades financieras a determinar, evaluar y administrar el riesgo ambiental y social en la financiación de proyectos para

Recuadro 5.1 *Evaluación social del Proyecto de Desarrollo y Crédito Agrícola de Azerbaiyán*

El Farm Privatization Project (Proyecto de Privatización de Parcelas), que otorgaba préstamos más flexibles y adaptables, se implementó con el fin de restaurar las zonas de cultivo de Azerbaiyán a sus anteriores niveles de productividad. El proyecto se concentró en el registro de propiedad, el desarrollo de mercados de la tierra y la provisión de crédito e información a mujeres y hombres rurales, especialmente de bajos ingresos.

El propósito de la evaluación social fue el de garantizar que la intervención propuesta estuviese basada en la propiedad (compromiso) del interesado, y que los beneficios previstos fueran socialmente aceptables. La información obtenida de la evaluación ayudó a los administradores del programa a desarrollar un proceso participativo de seguimiento y evaluación.

La primera fase de la evaluación social incluyó los componentes siguientes:

- Revisión de datos secundarios, incluyendo evaluaciones anteriores.
- Encuestas de hogares y mujeres en tres de las seis regiones, después de una evaluación cualitativa rápida.
- Entrevistas semiestructuradas de individuos (agricultores, administradores de fincas, trabajadores desempleados, dirigentes comunitarios, grupos de mujeres, asociaciones locales, técnicos, funcionarios del gobierno).
- Observación en el sitio por parte del personal (un integrante del equipo vivió con una familia de agricultores para efectuar una observación en el sitio, del impacto de la privatización de la parcela).
- Cinco grupos de discusión con grupos homogéneos de partes interesadas.
- Consultas con autores de políticas y administradores, e igualmente con ONG locales e internacionales.
- Discusiones con administradores anteriores de granjas del Estado y dirigentes comunitarios.
- Un seminario para las partes interesadas.

La evaluación se organizó alrededor de cuatro pilares:

- *Desarrollo social*: Principal interés en pobreza, género y exclusión social.
- *Instituciones*: La base de poder de las zonas rurales fue cambiando, por lo que se hizo difícil identificar las principales partes interesadas. Se efectuó investigación limitada sobre las organizaciones sociales y la falta de análisis de los impactos de la migración rural.
- *Participación*: Informe de confusión y ambigüedades en el proceso de reforma agraria. La distribución de la tierra redujo la pobreza y la influencia de los administradores anteriores, lo que contribuyó a apoderar la población rural. El acceso al crédito aumentó, pero las tasas de interés continuaron altas (15 a 18%).
- *Seguimiento y evaluación:* Se utilizaron indicadores de desempeño para controlar la ejecución. Los indicadores relacionaron los insumos y actividades del proyecto con medidas cuantificadas de los productos e impactos esperados.

En la evaluación se tuvieron también en cuenta los impactos en forma de una mayor productividad, mayores ingresos, reducción de pobreza y satisfacción de los participantes.

Fuente: Kudat y Ozbilgin, 1999.

Principios ecuatoriales: Enfoque industrial para determinar, evaluar y administrar el riesgo ambiental y social en la financiación de proyectos del sector privado.

el desarrollo con costos totales de capital de por lo menos US$10 millones (en 2006 US$50 millones). Su propósito es servir de base y marco para implementar procedimientos y estándares individuales, internos, ambientales y sociales para proyectos de desarrollo.

La norma ISO 14031

La International Organization for Standarization (ISO) desarrolló y mantiene estándares internacionales para la administración ambiental, conocidos como ISO 14031. El estándar, emitido por primera vez en 1999, son un proceso y herramientas de administración interna diseñados para suministrar a la administración información confiable y verificable en forma continua. Contribuye a determinar si el desempeño ambiental de una organización cumple los criterios fijados por su administración. La evaluación del desempeño ambiental y las auditorías ambientales ayudan a la administración a evaluar la condición de su desempeño ambiental y a identificar áreas de mejoramiento (ISO 1999).

La ISO 14031: Conjunto de estándares internacionales para la administración ambiental, desarrollado por la International Organization for Standarization.

La evaluación del desempeño ambiental se asiste estableciendo procesos para:

- Seleccionar indicadores.
- Recopilar y analizar datos.
- Evaluar información frente a criterios de desempeño ambiental (objetivos).
- Informar y comunicar.
- Revisar periódicamente y mejorar este proceso.

Estrategias de desarrollo sostenible: libro de recursos

Un libro de recursos publicado por la OCDE y el PNUD ofrece una guía flexible, sin prescripciones, sobre cómo desarrollar, evaluar e implementar estrategias nacionales para el desarrollo sostenible según los principios esbozados en los lineamientos sobre estrategias para el desarrollo sostenible (Dalal-Clayton y Ba, 2002). Contiene ideas y casos de estudio sobre las tareas principales en los procesos de estrategia. Sus lineamientos se dirigen a países, organizaciones e individuos interesados en el desarrollo sostenible a niveles nacional o local, y a organizaciones internacionales interesadas en el apoyo de dicho desarrollo.

Evaluación participativa

La evaluación participativa adopta la noción de participación del interesado a un nuevo nivel. Incluye responsabilidades compartidas para la planeación, implementación e información de la evaluación, haciendo participar a las partes interesadas en la definición

de preguntas para la evaluación, la recopilación y el análisis de datos, y el borrador y la revisión del informe.

Evaluación participativa: Evaluación en la cual las responsabilidades de planear, implementar e informar se comparten con las partes interesadas que contribuyen a definir preguntas para la evaluación, recopilar y analizar datos, y delinear y revisar el informe.

Paulmer (2005, p. 19) describe la evaluación participativa como:

> Una evaluación colectiva de un programa que realizan partes interesadas y beneficiarios. También se orientan a la acción, forman capacidad en las partes interesadas y facilitan la colaboración y la toma de decisiones compartida para utilizar más los efectos directos de la evaluación.

En la participación y los enfoques existen dos objetivos primordiales:

- La participación como un producto, en el que el acto de participación es un objetivo y uno de los indicadores del éxito.
- La participación es un proceso mediante el cual se obtiene un objetivo declarado.

Según Patton (1997), los principios básicos de la evaluación participativa son los siguientes:

- El proceso de evaluación incluye las destrezas de los participantes en la fijación de objetivos, el establecimiento de prioridades, la selección de preguntas, el análisis de datos y la toma de decisiones sobre los datos.
- Los participantes poseen (se comprometen con) la evaluación, ya que toman decisiones y sacan sus propias conclusiones.
- Los participantes aseguran que la evaluación se concentre en los métodos y efectos directos que consideren importantes.
- Las personas trabajan juntas, facilitando y promoviendo la unidad de grupo.
- Todos los aspectos de la evaluación son comprensibles y significativos para los participantes.
- Se valora mucho la responsabilidad propia.
- Los facilitadores actúan como recursos para el aprendizaje; los participantes actúan en la toma de decisiones y como evaluadores.

En la investigación convencional se han desarrollado y documentado lineamientos para técnicas específicas en un intento para aumentar la confiabilidad y la validez de sus datos. En la evaluación participativa no se opera según reglas claramente demarcadas heredadas de los expertos por los recopiladores de datos. Antes bien, los lineamientos para la recopilación de datos se desarrollan y evolucionan a través del consenso, la reflexión, el diálogo y la experiencia (Narayan, 1996).

El enfoque de evaluación participativa está recibiendo una mayor atención en el contexto del desarrollo. Se está utilizando más a menudo para proyectos de desarrollo, especialmente iniciativas basadas en la comunidad. La evaluación participativa es otro paso en el alejamiento del modelo de evaluación independiente o del evaluador como experto.

En el enfoque participativo se identifica y luego se involucra a las personas, agencias y organizaciones interesadas en el problema. Entre las personas se incluyen niños, mujeres y hombres en las comunidades, en especial los de grupos marginados. Se incluyen también personal de agencias, autores de políticas y todos los afectados por las decisiones tomadas en el proceso de investigación participativa (Narayan, 1996).

En la evaluación participativa se les solicita a las partes interesadas mantener un diario o una bitácora de sus experiencias con una intervención. Pueden auxiliar con entrevistas a otros integrantes de la comunidad y también participar en el análisis de los datos, la interpretación de los hallazgos y hacer o desarrollar recomendaciones.

Las decisiones de planeación, por ejemplo identificación de preguntas, medidas y estrategias de recopilación de datos, se toman junto con los participantes. Se trata de un proceso conjunto y no de un proceso tradicional descendente (cuadro 5.2).

El enfoque participativo usualmente incrementa la credibilidad de los efectos directos de la evaluación por parte del personal del programa, y también la probabilidad de que se utilicen los efectos directos. Los defensores de la evaluación participativa la ven como una herramienta para apoderar a los participantes y aumentar la capacidad local para ocuparse en el proceso de desarrollo.

La evaluación participativa plantea retos considerables. Implica mayores costos transaccionales que una evaluación tradicional, ya que sostener reuniones y asegurarse de que todos entienden lo que se espera consume mucho tiempo y requiere muchas destrezas. Los grupos suelen pasar por un proceso en el que se reconcilian las diferencias y se desarrollan las normas de grupo antes de que el grupo se concentre en alcanzar las tareas a mano. Este proceso de dinámica de grupos se conoce algunas veces como "formación, tormenta, normas y acción". Después de la formación, es natural llegar a un período de conflicto (tormenta). Si el grupo resuelve estos conflictos, establecerá acuerdos más específicos sobre cómo trabajarán sus miembros conjuntamente (normas). Una vez establecidos los acuerdos, pasarán a realizar las tareas a mano (acción).

Puede haber también retos para crear un equipo igualitario en una cultura en la que distintos miembros tienen una posición diferente en la comunidad. El evaluador que desea realizar una evaluación participativa debe tener destrezas de facilitación, cola-

Cuadro 5.2 *Características de las técnicas de evaluación participativa y tradicional*

Participativa	*Tradicional*
• Enfoque y propiedad de los participantes	• Enfoque y propiedad de los donantes
• Enfoque en el aprendizaje	• Enfoque en rendición de cuentas y juicio
• Diseño flexible	• Diseño predeterminado
• Métodos más informales	• Métodos formales
• Personas externas como facilitadores	• Personas externas como evaluadores

Fuente: los autores.

boración y gestión de conflictos (o hacer que alguien que tenga esas destrezas asuma el liderazgo). Él o ella deben tener también la capacidad de ofrecer capacitación del tipo justo a tiempo sobre destrezas básicas y técnicas asociadas con la evaluación y los procesos grupales inherentes a la participación.

Las personas capacitadas en evaluación tradicional y que la realizan pueden preocuparse porque una evaluación participativa no sea objetiva. Existe el riesgo de que aquellos que se encuentran más cerca de la intervención no puedan ver lo que está sucediendo realmente, sino lo que esperan ver. La evaluación puede, en realidad, verse "capturada" y perder objetividad. Los participantes pueden estar temerosos de suscitar opiniones negativas por temor de que otros integrantes del grupo puedan desterrarlos o terminar la intervención, lo que produciría pérdidas de dinero para la comunidad, o que nunca puedan lograr que la organización para el desarrollo trabaje con ellos nuevamente. Aproximarse a evaluaciones participativas desde una perspectiva de aprendizaje puede contribuir a reducir estos temores y los evaluadores deberían analizar el grado al cual pueda verse comprometida la credibilidad (en opinión de los agentes externos) al escoger un enfoque de evaluación participativa y no una evaluación independiente.

Gariba (1998) describió cómo la palabra *evaluación* causa reacciones mixtas entre las entidades donantes y quienes implementan la intervención. Los donantes se preocupan sobre cómo afectará la evaluación un proyecto (esto es, ocasionar que se extienda o termine). Los ejecutores del proyecto temen que una evaluación denigre su enfoque de administración de proyecto.

Gariba describe cómo puede ser la evaluación participativa una forma exitosa y sistemática de aprender de la experiencia. Con la evaluación participativa, los socios en la intervención para el desarrollo pueden adquirir enseñanzas de su interacción y adoptar medidas correctivas para mejorar la eficacia o la eficiencia de sus futuras actividades.

Gariba (1998) describe tres elementos críticos de la evaluación participativa:

- *La evaluación como herramienta de aprendizaje*: Este principio conforma el principal paradigma de elección. El propósito no es investigar sino crear una oportunidad para que todas las partes interesadas, incluidas las entidades donantes, aprendan de su papel en el ejercicio de la intervención para el desarrollo.
- *La evaluación como parte del proceso de desarrollo*: La actividad de la evaluación no es discreta y separable del proceso de desarrollo mismo. Los efectos directos y herramientas correspondientes se convierten en herramientas del cambio en lugar de quedar como informes históricos.
- *La evaluación como asociación de responsabilidades compartidas*: En la metodología de evaluación de impacto participativa, todos los actores tienen más o menos igual peso (lo que está en agudo contraste con la tendencia de los evaluadores de establecer el síndrome de "nosotros" los profesionales y "ellos" los actores y beneficiarios del proyecto). El evaluador se transforma de investigador en promotor y participante.

001El camino.indd 181 08/11/2010 03:27:47 p.m.

De acuerdo con la Canadian International Development Agency Guide (CIDA, 2004), si las partes interesadas participan en el desarrollo de los efectos directos, es más probable que contribuyan a implementar la intervención. La CIDA cree que la evaluación participativa también

- Forma responsabilidad dentro de las comunidades.
- Ofrece una orientación más realista a la evaluación.
- Incrementa la cooperación.
- Apodera a los participantes locales al incluirlos en el proceso de evaluación.

En el recuadro 5.2 se ilustra una técnica utilizada en la evaluación participativa.

Cartografía de los efectos directos

El Centro Internacional de Investigaciones para el Desarrollo (CIID) desarrolló un enfoque innovador para la evaluación. Su enfoque de cartografía de efectos directos no intenta reemplazar formas más tradicionales de evaluación, sino complementarlas concentrándose en el cambio en el comportamiento relacionado.[1]

La cartografía de efectos directos se centra en un tipo específico de resultado: el cambio en el comportamiento. El foco está en los efectos directos y no en el logro de impactos en el desarrollo, que se consideran demasiado "secundarios" y que reflejan muchos esfuerzos e intervenciones. Según argumenta el CIID, no tiene sentido intentar evaluar con precisión la contribución de cualquier organización al impacto. Más bien, la cartografía de efectos directos busca fijarse en el comportamiento resultante de esfuerzos múltiples, a fin de mejorar el desempeño de los proyectos, programas y políticas.

Cartografía de los efectos directos: Es la cartografía de cambio en el comportamiento.

Con la cartografía de efectos directos, se identifica a los socios demarcados: individuos, grupos y organizaciones que interactúan con proyectos, programas y políticas. En la cartografía de efectos directos se asume que los socios demarcados controlan el cambio y se asume también que su papel como agente externo es lo que les proporciona acceso a nuevos recursos, ideas u oportunidades por un determinado tiempo. Los programas más exitosos, de acuerdo con los defensores de la cartografía de efectos directos, son aquellos que transfieren poder y responsabilidad a las personas que actúan dentro del proyecto o programa.

El foco de la cartografía de efectos directos es la gente, lo que representa una desviación de evaluar el impacto en el desarrollo de un proyecto o un programa, hacia describir cambios en la forma como se comporta la gente mediante las acciones ya sea individualmente o dentro de grupos o de organizaciones. La cartografía de efectos directos proporciona una forma de modelar lo que se intenta hacer con un programa.

1 Gran parte de la información de esta sección se adaptó de Earl, Carden y Smutylo (2001).

Recuadro 5.2 *Generar confianza a través de la evaluación participativa*

En Marruecos una facilitadora de grupos generó confianza en las mujeres, pidiéndoles formar un círculo tomadas de la mano. Luego les pidió enredarse entre ellas sin soltarse las manos. Un actor "externo" instruyó entonces al grupo a desenredarse, ejercicio que tomó seis minutos. Luego se solicitó al grupo repetir el ejercicio y esta vez las mujeres se desenredaron en 10 segundos.

La facilitadora le ayudó al grupo a llegar a una conclusión sobre los papeles de los "externos": éstos generalmente funcionan mejor como facilitadores y catalizadores que como líderes. La población local sabe encontrar mejor las soluciones a sus problemas, pues los vive.

Fuente: Banco Mundial, 1996.

Difiere de los modelos lógicos más tradicionales porque reconoce que diferentes socios demarcados operan dentro de distintos sistemas lógicos y de responsabilidad.

La cartografía de efectos directos ofrece un método para controlar los cambios en los socios demarcados y en el programa visto como una organización. Alienta al programa a evaluar regularmente cómo puede mejorar su desempeño. La cartografía de efectos directos puede utilizarse también como herramienta de evaluación al final del programa, cuando el propósito de la evaluación es estudiar el programa como un todo.

Los partidarios de la cartografía de efectos directos creen que muchos programas, especialmente los que se centran en la formación de capacidad, pueden mejorar la planeación y la evaluación de su contribución al desarrollo, enfocándose en el comportamiento. Por ejemplo, un objetivo de un programa puede ser proveer a las comunidades acceso a agua más limpia, instalando filtros de purificación. Con un método tradicional se podrían evaluar los efectos directos, contando el número de filtros instalados y midiendo los cambios en el nivel de contaminantes del agua antes y después de instalados. Un enfoque de cartografía de efectos directos se centraría en el comportamiento y empezaría con la premisa de que el agua no seguirá limpia si las personas no mantienen su calidad a través del tiempo. Los efectos directos del programa se evaluarían entonces enfocándose en el comportamiento de los responsables de la pureza del agua; específicamente, cambios en su adquisición y uso de las herramientas, las destrezas y el conocimiento apropiados. Con la cartografía de efectos directos se evaluaría cómo controlan las personas los niveles de contaminantes, cómo cambian los filtros o acuden a expertos cuando se requiera.

Una canción sobre la cartografía de efectos directos, escrita por Terry Smutylo, anterior director de evaluación del Centro Internacional de Investigaciones para el Desarrollo, resume algunos de los problemas que busca resolver. (En la lista de sitios web al final de este capítulo se presenta el vínculo de una grabación de Terry Smutylo cantando "The Output Outcome Downstream Impact Blues".)

001El camino.indd 183 08/11/2010 03:27:48 p.m.

The Output Outcome Downstream Impact Blues
Productos, efectos directos, impactos: ¿para quién, por quién, quién lo dice?

Coro

No busques *impacto* con atribución (4 veces)
Pues hay por ahí una palabreja que se usa demasiado
En los programas de desarrollo hay propensión al abuso
Es ya una obsesión ahora que estamos todos en el acto
Pues la supervivencia depende de ese elusivo *impacto*

Estribillo I

Pues es *impacto* en todas partes, *impacto* en todo tiempo
Puedes bien encontrarlo a la vuelta de la esquina o mucho más lejos
Pero si sucede en la forma no escogida
Obtienes esos *blues de impacto producto resultado secundario.*

Ahora cuando los donantes buscan el *impacto* que realmente quieren ver
Es una pequeña imagen de su fantasía
Esto es algo que un buen evaluador nunca haría:
Utilizar una palabra como *impacto* sin pensarlo bien.

Pero ya los donantes dicen que es un hecho
Sal y muéstranos tu *impacto*
Debes cambiar la vida de la gente y ayudarnos a obtener el crédito
O la próxima vez que necesites fondos… mmm, mmm
Quizá no los consigas.

Estribillo I

Pues es *impacto* en todas partes, *impacto* en todo tiempo
Puedes bien encontrarlo a la vuelta de la esquina o mucho más lejos
Pero si sucede en la forma no escogida
Obtienes esos *blues de impacto producto resultado secundario.*

Bien están los receptores dispuestos siempre a complacer
Cuando enviamos nuestros evaluadores al exterior
Para buscar indicadores de *impacto* mesurable
Sorprendiendo a los donantes con lo que traen.

Bueno, *impacto* lo encuentran cuando éste ocurre
Por muchos factores y no estamos seguros
De qué podemos atribuir a quién
Porque *impacto* es producto de lo que muchos hacen.

Estribillo II

Pues es *impacto* en todas partes, *impacto* en todo tiempo
Puedes bien encontrarlo a la vuelta de la esquina o mucho más lejos

Pero si buscas atribución nunca perderás
esos *blues de impacto producto resultado secundario.*

Así los donantes despiertan de tu sueño imposible
Tú dejas caer tus fondos largo trecho río arriba
Entonces en las aguas ellos fluyen, se mezclan y fusionan
Entonces, ¿cómo puedes tener crédito por lo que al final resulta?

Estribillo II

Pues es *impacto* en todas partes, *impacto* en todo tiempo
Puedes bien encontrarlo a la vuelta de la esquina o mucho más lejos
Pero si buscas atribución nunca perderás
esos *blues de impacto producto resultado secundario.*

Coro (4 veces, luego se desvanece)

Evaluación rápida

Evaluación rápida: Enfoque de evaluación sistemática, semiestructurada realizada en el campo, usualmente por un equipo de evaluadores.

Las evaluaciones rápidas suplen la necesidad de valoraciones rápidas y de bajo costo. Son especialmente útiles en el contexto de la evaluación en los países en desarrollo, donde las restricciones de tiempo y otras (como falta de datos básicos, bajos niveles de alfabetización que hacen necesarios los cuestionarios administrados y limitados presupuestos para evaluación, para mencionar unos pocos) dificultan a menudo hacer una evaluación más completa.

Aunque no hay una definición fija de una evaluación sistemática, generalmente se la describe como un puente entre la recopilación formal e informal de datos, un "enfoque razonablemente rápido y limpio" en lugar de uno "rápido y sucio". Puede describirse como un enfoque de evaluación sistemática, semiestructurada realizada en el campo, usualmente por un equipo de evaluadores. Idealmente, el tema es diverso, de modo que refleja varias perspectivas.

La evaluación rápida tiene un mejor uso cuando se examinan procesos y no efectos directos o impactos. Por lo general, busca recopilar sólo la información más esencial –el "hay que saber" en lugar del "es bueno saber"– obtenida mediante métodos cuantitativos y cualitativos. Su orientación básica en la evaluación para el desarrollo es "buscar para comprender", porque es más probable que un enfoque no crítico produzca como respuesta conversaciones abiertas y honestas.

Se efectúan visitas al sitio, porque la observación de una intervención dentro de su entorno puede ofrecer indicaciones sobre su buen funcionamiento. Una tarea básica es la de identificar para entrevistas a personas que tengan una variedad de experiencias y perspectivas, en especial aquellas con mayor probabilidad de ser omitidas en una evaluación. Un grupo pequeño pero altamente diverso de informantes clave puede ser muy eficaz para obtener una opinión integral de la situación. La habilidad de escuchar es esencial.

En la evaluación rápida se pueden utilizar los mismos métodos de recopilación y análisis de datos que en cualquier otro tipo de evaluación, y la diferencia suele ser sólo de alcance. Comúnmente las evaluaciones rápidas son de corto alcance y en ellas se contacta a unas pocas personas en unas cuantas ubicaciones. Debe utilizarse más de una fuente de información, porque varias fuentes aumentan la credibilidad, reducen los prejuicios y ofrecen una perspectiva integral. Los datos de las entrevistas de los informantes deben, por tanto, complementarse con la información de informes y estudios anteriores, registros y documentos, lo mismo que con los datos recopilados mediante la observación, grupos de discusión, entrevistas grupales y encuestas. Cuanto más consistente sea la información de estas fuentes, los hallazgos generarán mayor confianza.

Según la medida en que se utilicen los métodos cualitativos, serán esenciales destrezas fuertes en la toma de anotaciones, y algo que ayuda mucho es que el evaluador mantenga un diario para anotaciones de observaciones, temas, corazonadas, interpretaciones y cualquier incidente que suceda en la visita de campo. Estas anotaciones deben compartirse con los demás miembros del equipo para identificar problemas comunes.

Al efectuar una evaluación rápida, los evaluadores deben tener presentes las indicaciones siguientes:

- Revisar datos secundarios antes de hacer las visitas de campo.
- Una vez en el campo, observar, escuchar, preguntar y registrar.
- Mantener buenas anotaciones durante todo el proceso; las buenas anotaciones no solamente son esenciales para el informe, sino que también contribuyen a que tenga sentido la información recopilada por los distintos integrantes del equipo.

Las estrategias y enseñanzas aprendidas al realizar evaluaciones rápidas indican que antes de emprenderla debe considerarse lo siguiente:

- Crear un equipo diverso y multidisciplinario integrado por hombres y mujeres.
- En lo posible, reclutar personas de confianza (familiarizadas con la intervención y la zona local) y externas (quienes verán las cosas en una forma fresca).
- Utilizar equipos pequeños en lugar de grandes para maximizar las interacciones.
- Dividir el tiempo entre la recopilación de los datos y darles sentido.
- Estar dispuestos a ir donde sea necesario (campos, mercados, zonas apartadas de las vías principales), y no sólo donde sea conveniente.
- Mantener la flexibilidad y la adaptabilidad, pues una nueva información puede modificar el plan de evaluación (FAO, 1997).

Síntesis de evaluación y metaevaluación

Una síntesis de evaluación constituye un enfoque útil cuando ya se han efectuado muchas evaluaciones de los efectos directos de intervenciones similares, y capacita al evaluador para examinar intervenciones dirigidas a un problema o asunto similar a fin

de identificar características comunes. Es útil cuando con la evaluación se busca averiguar la eficacia general de una intervención.

Síntesis de evaluación: Enfoque en donde un evaluador examina distintas intervenciones dirigidas a un problema o un asunto similar para identificar características comunes.

Chelimsky y Morra (1984) extendieron el método a un contexto más amplio de políticas cuando aplicaron la técnica a evaluaciones de intervenciones predominantemente cualitativas y cuantitativas. Antes de que ellos hubiesen hecho esto, la síntesis de evaluación se había utilizado con evaluaciones que se prestaban al análisis predominantemente cuantitativo; por ejemplo, los efectos de los programas de nutrición para mujeres jóvenes en el peso y en la mortalidad de los hijos al nacer, o los efectos del tamaño de la clase en el desempeño educativo de los estudiantes.

El concepto subyacente en la síntesis de evaluación es que mientras las evaluaciones individuales pueden aportar información útil acerca de una intervención específica, cada una es a menudo demasiado cualificada y específica de su contexto como para permitir un enunciado general sobre el impacto de la intervención. Sin embargo, cuando se combinan los efectos directos de muchos estudios, se pueden hacer enunciados generales sobre el impacto de una intervención (y aun de una política).

Una ventaja de la síntesis de evaluación es que utiliza la investigación disponible, por lo que hacerla se hace menos costosa que otros tipos de evaluación. Otra ventaja es que crea una mayor base para evaluar el impacto de una intervención, aumentando la confianza al hacer enunciados generales sobre el impacto. Los problemas están en localizar los estudios pertinentes, publicados y no publicados; determinar la calidad de cada evaluación y, cuando sea aplicable, obtener los conjuntos de datos para las combinaciones en el análisis secundario.

Una síntesis de evaluación debería incluir lo siguiente:

- Procedimientos claros para identificar las evaluaciones y definir su alcance.
- Criterios transparentes de calidad para tomar decisiones sobre la inclusión o la exclusión de evaluaciones en la síntesis.
- Procedimientos para aplicar los criterios de calidad (a menudo realizados por un panel y no por un individuo).
- Citas de todas las evaluaciones revisadas.
- Descripciones sumarias de cada evaluación incluida en la síntesis y los hallazgos sobre los problemas o variables de interés.
- Vacíos o limitaciones de la síntesis.

A veces sólo se utilizan algunos elementos de una síntesis de evaluación. Por ejemplo, en el caso descrito en el recuadro 5.3 no parece ser que los informes hayan sido tamizados por la calidad, lo que genera alguna duda sobre los hallazgos de la síntesis.

Los términos *síntesis de evaluación* y *metaevaluación* se utilizan en ocasiones en forma intercambiable en la literatura sobre la evaluación, pero aquí distinguimos entre

Metaevaluación: Revisión experta de una o más evaluaciones frente a estándares de calidad profesional.

ambos. En la forma como aquí se utiliza, la síntesis de evaluación se refiere a un sumario analítico de efectos directos de las evaluaciones que cumplen estándares mínimos de calidad. En contraste, la metaevaluación se refiere a una revisión experta de una o más evaluaciones frente a estándares de calidad profesional.

Otros enfoques

Otros enfoques, teorías y modelos se utilizan asimismo en las evaluaciones para el desarrollo, la mayor parte de ellas constituyen variaciones sobre el tema de la evaluación participativa. Los lectores encontrarán elementos de algunos modelos en la evaluación basada en efectos directos que se promueve en este volumen, como la dependencia de especificar y someter a prueba una teoría del cambio, y centrarse en el uso desde el principio, y seguir el proceso de evaluación identificando e incluyendo a las partes interesadas clave.

En esta sección se describen los enfoques siguientes:

- Evaluación centrada en el uso.
- Evaluación para el empoderamiento.
- Evaluación realista.
- Evaluación incluyente.
- Evaluación sobre los beneficiarios.
- Evaluación horizontal.

Recuadro 5.3 *Uso de una síntesis para medir los beneficios ambientales*

En 2000, el Department for International Development (DfID) del Reino Unido publicó un estudio de síntesis de evaluación sobre el medio ambiente. El DfID había administrado con éxito un portafolio sustancial de proyectos ambientales en los años noventa, pero tenía la sensación de que los beneficios ambientales "se asumían en forma general en lugar de examinarse en forma crítica" (p. 1). Se comisionó el estudio de síntesis ambiental a fin de examinar 49 proyectos apoyados por el DfID en cinco países, y evaluar la implementación y el impacto del apoyo bilateral a proyectos para el mejoramiento y la protección del medio ambiente. Los proyectos no fueron primordialmente ambientales, pero cubrían una amplia variedad de intervenciones ambientales (eficiencia energética, industria, silvicultura, biodiversidad, agricultura y mejoramiento urbano).

Después de examinar someramente los 49 estudios, los evaluadores concluyeron que había "un vacío entre la alta prioridad política asignada por el DfID a los problemas ambientales… y lo que se había llevado a cabo realmente en términos de impacto ambiental positivo".

Fuente: DfID, 2000, 1.

Evaluación centrada en el uso

En la evaluación centrada en el uso se propone que una evaluación debe juzgarse por su utilidad y por cómo se utiliza realmente. Patton (2002, p. 173), cuyo principal libro sobre la evaluación centrada en el uso ya está en su cuarta edición (2008), la describe como empezar con la "identificación y organización de personas específicas que toman las decisiones y que utilizan la información (no la audiencia vaga y pasiva), quienes utilizarán la información que produce la evaluación". Se trata de la evaluación centrada en el uso previsto por parte de los usuarios previstos, quienes son aquellos que tienen la responsabilidad de aplicar los hallazgos de la evaluación y llevar a la práctica sus recomendaciones. La evaluación centrada en el uso auxilia a los usuarios previstos primordiales a seleccionar el modelo de evaluación más apropiado, el contenido y los métodos para su situación particular.

Evaluación centrada en el uso: Es la evaluación que se juzga según su utilidad y en cómo se utiliza realmente.

Evaluación para el empoderamiento

La evaluación para el empoderamiento es el uso de conceptos, técnicas y hallazgos de la evaluación para fomentar el mejoramiento y la autodeterminación (Fetterman, Kaftarian y Wandersman, 1996). Reconoce un profundo respeto por la capacidad de las personas de crear conocimiento sobre su propia experiencia y encontrar soluciones a los problemas que encuentran. Al ayudarles a las personas a lograr sus objetivos como individuos y miembros de una comunidad y a mejorar su vida, el apoderamiento puede crear un sentido de bienestar y crecimiento positivo (Fetterman y Wandersman, 2004).

Evaluación para el empoderamiento: Uso de conceptos, técnicas y hallazgos de la evaluación para fomentar el mejoramiento y la autodeterminación.

Fetterman y Wandersman (2004) describen el papel de un evaluador para el empoderamiento como el de un "amigo crítico", y abogan porque los integrantes de la comunidad continúen a cargo de la evaluación; el evaluador debe tener el papel de facilitador, influyendo en la evaluación, pero no controlándola.

La evaluación para el empoderamiento comparte algunas características de la evaluación centrada en el uso. Ambos enfoques están diseñados para ser provechosos, constructivos y útiles en cada una de las etapas de la evaluación (Fetterman y Wandersman, 2004). Pero Patton (1997) indica su reopinión de que mientras la evaluación para el empoderamiento traslapa los enfoques participativos concernientes a la propiedad, la pertinencia y la formación de capacidad, el foco que la define es fomentar la autodeterminación. La evaluación para el empoderamiento va más allá de la mayor parte de la evaluación participativa, en el sentido de que el evaluador-facilitador es un defensor de los no apoderados y promueve una agenda de cambio político.

Al describir la diferencia entre las evaluaciones para el empoderamiento y participativa, Alkin y Christie (2004, p. 56) afirman:

> Como la evaluación participativa surge de un marco de utilización, el objetivo de la evaluación participativa es la mayor utilización a través de estas actividades [diseño, implementación, análisis e interpretación] en forma opuesta al apoderamiento de aquellos que han sido oprimidos, que es de naturaleza política o de emancipación.

Evaluación realista

Pawson y Tilley (2004, p. 1) describen la evaluación realista como una "especie de evaluación dirigida por la teoría", y la relacionan con la teoría del cambio y la teoría programática. Para construir la teoría, debe comprenderse el contexto general en donde tiene lugar la intervención, cómo funciona la intervención en el contexto particular y si es que funciona.

Según Pawson y Tilley (2004, p. 22), la evaluación realista proporciona un "marco coherente y consistente" para la manera como las evaluaciones se ocupan de los programas. Reconoce la importancia de las partes interesadas en el desarrollo y la ejecución del programa, pero dirige un curso entre pasar por alto las partes interesadas (por sus sesgos de interés propio) y considerarlos como omniscientes e infalibles (por su conocimiento interior). Las partes interesadas son tratadas como expertos falibles cuyo conocimiento se debe formalizar y poner a prueba (Pawson y Tilley, 2004).

Evaluación realista: Dirigida por la teoría que presenta un marco coherente y consistente para la manera como las evaluaciones se ocupan de los programas, trata a las partes interesadas como expertos falibles y utiliza otros enfoques para la evaluación.

La evaluación realista se deriva de una amplia variedad de enfoques de investigación y evaluación y usa partes o todos ellos. La evaluación realista puede ser cualitativa o cuantitativa, y a menudo combina ambos métodos, pero no establece una relación causal entre una intervención y un resultado por experimentación. La relación causal se establece delineando la teoría detrás de cómo funciona la intervención. Los pasos se ejecutan en una secuencia lógica para probar la hipótesis y determinar lo que podría funcionar, para quién y en qué circunstancias. Pawson y Tilley (2004) indican que la evaluación realista puede ser problemática porque ninguna fórmula simple ofrece recetas simples para presentar los hallazgos.

Evaluación incluyente

Se centra en incluir a los miembros menos favorecidos de una población como parte de una investigación sistemática del mérito o valor de un proyecto, un programa o una política. La evaluación incluyente se basa en datos, pero los datos se generan de las partes interesadas menos favorecidas, aquellas que tradicionalmente no han tenido representación suficiente. En una evaluación incluyente no se incluye a quienes tradicionalmente se incluye en las evaluaciones (Mertens, 1999). A semejanza de la evaluación para el empoderamiento, la evaluación incluyente es un paradigma transformacional.

Evaluación incluyente: Evaluación en la que se incluye a los miembros menos favorecidos de una población.

Una evaluación incluyente haría preguntas como estas:

- ¿Cuáles son las diferencias importantes dentro de la población que se va a atender?
- ¿Cómo se prestan los servicios dentro de los distintos subgrupos?
- ¿Cuáles son los valores subyacentes a la distribución de los servicios?

Evaluación sobre los beneficiarios

Evaluación sobre los beneficiarios: Herramienta utilizada para mejorar el impacto de las operaciones para el desarrollo, obteniendo la opinión de los beneficiarios previstos con respecto a una intervención planeada o en curso.

Es "una herramienta de investigación cualitativa utilizada para mejorar el impacto de las operaciones para el desarrollo, obteniendo la opinión de los beneficiarios previstos con respecto a una intervención planeada o en curso" (Salmen, 1999, p. 1). A semejanza de la evaluación incluyente, en este enfoque se persigue incluir grupos que a menudo son pasados por alto. Este enfoque centrado en el proyecto fue desarrollado por el Banco Mundial, a fines de los años ochenta, para complementar su enfoque de evaluación más técnico y financiero, y generalmente se ha aplicado a proyectos con componente de prestación de servicios.

La evaluación sobre los beneficiarios incluye al cliente final, los beneficiarios del proyecto. El fundamento es que una mayor participación de los beneficiarios para conformar el diseño del programa, suministrar retroalimentación de seguimiento y presentar sus opiniones sobre los impactos, aumenta su propiedad haciéndolos actores clave para producir los cambios necesarios y deseados en su propio desarrollo.

El objetivo de la evaluación sobre los beneficiarios es evaluar el valor de una actividad según lo perciben los beneficiarios del proyecto, e integrar esos hallazgos en las actividades del proyecto. La evaluación sobre los beneficiarios tiene un papel central en la evaluación social al ayudar a conectar la cultura con la toma de decisiones (Salmen, 1999).

Evaluación horizontal

Evaluación horizontal: La que combina un proceso interno de evaluación con una revisión externa de pares.

Es un enfoque de evaluación relativamente nuevo que combina un proceso interno de evaluación con una revisión externa de pares. La combinación se diseñó para neutralizar "las relaciones de poder desequilibradas predominantes en las evaluaciones externas tradicionales, creando una atmósfera más favorable para el aprendizaje y el mejoramiento posterior de los programas" (Thiele y otros, 2006, p. 1).

La evaluación horizontal se ha utilizado con mucha frecuencia para el aprendizaje y el mejoramiento de las metodologías de investigación y desarrollo que están en proceso. El enfoque se ha utilizado en un programa regional andino que desarrolla nuevas tecnologías de investigación y desarrollo, y en Uganda para evaluar el enfoque de cadena de mercado participativa (Thiele y otros, 2006, p. 1).

La clave para el enfoque de evaluación horizontal es que hay dos grupos separados de partes interesadas. El primero está conformado por participantes locales, que presentan y critican el proceso bajo investigación y hacen recomendaciones sobre cómo mejorarlo. El segundo está conformado por los visitantes (pares de otras organizaciones o proyectos que trabajan en asuntos similares), que evalúan el proceso, identifican las fortalezas y debilidades y hacen sugerencias para el mejoramiento (Thiele y otros, 2006). Un componente de la evaluación horizontal es un taller que permite reunirse a los dos grupos.

Resumen

Un enfoque de evaluación es una forma de considerar o de conceptuar una evaluación en una forma general. A menudo incorporan una filosofía y un conjunto de valores. Algunos enfoques han estado en uso durante años, mientras otros se han desarrollado relativamente hace poco tiempo, o se han aplicado más recientemente a la evaluación para el desarrollo. Las características básicas de los enfoques de evaluación se resumen en el cuadro 5.3.

Los enfoques descritos se han utilizado principalmente a nivel de la intervención. En el capítulo 11 se tratan las evaluaciones complejas de intervenciones complejas para el desarrollo, que se están convirtiendo ya en la norma.

001El camino.indd 192 08/11/2010 03:27:48 p.m.

001El camino.indd 193

Cuadro 5.3 *Propósito, características, fortalezas y problemas de los distintos enfoques de evaluación*

Enfoque de evaluación	*Propósito / Filosofía*	*Características / Actividades*	*Fortalezas*	*Problemas*
Prospectiva	Revisa la intervención antes de su inicio Responde preguntas hacia el futuro sobre el éxito probable	Ofrece un análisis conceptual del proyecto, programa o política Revisa las evaluaciones efectuadas de programas similares para identificar enseñanzas y problemas Predice el éxito o el fracaso y sugiere formas de fortalecer la intervención propuesta si sigue adelante	El uso del análisis secundario mantiene bajos los costos de la evaluación Puede tratar y resolver problemas, y fortalecer un programa antes de su inicio	Puede contar con cuerpo de informes de evaluación sobre programas que guardan estrecha relación
Evaluación de la evaluabilidad	Determina si la evaluación sería útil y factible Facilita la comunicación entre evaluadores y partes interesadas Clarifica el propósito del estudio formativo o las funciones como herramienta de planeación	Revisa materiales y datos disponibles Revisa el conocimiento compartido de las metas y objetivos del programa Entrevista a los administradores de la intervención, las partes interesadas y personas y organizaciones que puedan implementar las recomendaciones Desarrolla el modelo de la teoría del cambio	Puede conducir a una evaluación más realista y apropiada, y a programas mejor diseñados Puede construir apoyo de las partes interesadas y mejorar la comprensión de metas y objetivos Puede impedir gastos en evaluaciones sin capacidad de responder preguntas por vacíos en los datos	Puede retrasar innecesariamente la evaluación si se aplica a todos los programas antes de efectuarse la evaluación
Basado en objetivos	Mide el cumplimiento de las metas y los objetivos Sirve de base para la mayoría de los sistemas de evaluación de proyectos de los donantes	Identifica metas y objetivos Evalúa si la intervención logra las metas y objetivos (evaluación normativa)	Puede simplificar la evaluación en metodología clara, comparando realidad y estándares	Puede omitir importantes efectos no declarados explícitamente como metas u objetivos
Sin tener en cuenta las metas	Abre la opción de recopilar datos sobre los efectos y la eficacia de los programas sin estar restringida por un foco estrecho de metas y objetivos declarados	Busca impedir a los evaluadores ser capturados por las metas y los objetivos del programa, limitando el contacto con el director y el personal del programa	Aumenta la probabilidad de que se noten efectos laterales imprevistos	Permite un contacto limitado con el personal del programa; puede ser difícil para los evaluadores evitar conocer las metas y objetivos del programa

(Continúa en la página siguiente)

08/11/2010 03:27:48 p.m.

001El camino.indd 194

(Continuación cuadro 5.3)

Enfoque de evaluación	*Propósito / Filosofía*	*Características / Actividades*	*Fortalezas*	*Problemas*
Evaluación en sitios múltiples	Investiga intervenciones con ejecución estándar en todas las ubicaciones o variaciones planeadas, para determinar las condiciones bajo las cuales el programa logre de mejor forma sus metas y objetivos Debe capturar el clima en que operan las intervenciones, al igual que las diferencias sistemáticas culturales, geográficas, económicas, de tamaño u otras, que puedan afectar las variaciones	Forma participación de las partes interesadas Recopila información más profunda que algunos otros enfoques Selecciona sitios para representar ciertas características que puedan resultar en diferencias sistemáticas en la implementación de la intervención y los efectos directos Describe y compara intervenciones dentro de su contexto para ofrecer varias enseñanzas y estrategias para tratar con una variedad de situaciones	Puede llevar a menudo a la identificación de buenas prácticas Obtiene hallazgos más fuertes que la evaluación de una sola intervención en un solo sitio Puede tratar análisis dentro de cada sitio y entre ellos	Requiere recopilación de datos estandarizados Requiere personal bien capacitado, acceso a todos los sitios y suficiente información con anticipación para diseñar instrumentos de recopilación de datos Puede medir hallazgos muy específicos al contexto por sitio, que no son representativos
Evaluación de conjunto	Examina grupos de intervenciones similares o relacionadas que juntas representan un conjunto de intervenciones Examina el "grupo" de intervenciones en uno o múltiples entornos	Forma participación de las partes interesadas Resta énfasis a la generalización y replicación Considera positiva la variación Utiliza múltiples casos de estudio, con información compartida entre ellos	Se centra en el aprendizaje y no en sacar conclusiones generales sobre la calidad o el valor del programa	Muestra resultados agregados y no sobre sitios individuales
Evaluación social	Examina varias estructuras, procesos y cambios sociales dentro de un grupo o una comunidad Actúa como instrumento principal para asegurar que se examinen los impactos sociales de las intervenciones para el desarrollo	Investiga la consistencia entre los objetivos de la intervención y las necesidades, los intereses y capacidades de las partes interesadas Trata los efectos de los factores sociales y culturales sobre la capacidad de las partes interesadas de participar en, o beneficiarse de, la intervención	Identifica impactos adversos sociales potenciales, de modo que puedan ponerse a funcionar estrategias de mitigación	Puede perder el foco sobre las interacciones con problemas ambientales, atendiendo sólo parte de la ecuación

(Continúa en la página siguiente)

08/11/2010 03:27:48 p.m.

001El camino.indd 195

(Continuación cuadro 5.3)

Enfoque de evaluación	*Propósito / Filosofía*	*Características / Actividades*	*Fortalezas*	*Problemas*
		Investiga el impacto diferencial de la intervención en subconjuntos de partes interesadas		
Evaluación ambiental y social	Evalúa el logro de objetivos ambientales y sociales	Se centra en los aspectos sociales y ambientales de todas las intervenciones, no sólo de aquellas que se centran únicamente en el entorno o las desigualdades sociales Es útil cuando la evaluación ambiental y social constituye el único propósito de la evaluación	Enfatiza en el dominio de la importancia de los aspectos sociales y ambientales de las intervenciones	Puede requerir pericia técnica
Evaluación participativa	Comparte con las partes interesadas la responsabilidad por la planeación, implementación e información sobre la evaluación Examina la participación como un producto, donde el acto de participación es un objetivo y uno de los indicadores de éxito Considera la participación también como un proceso por el cual alcanzar el objetivo deseado	Incluye a los participantes en la fijación de metas, el establecimiento de prioridades, la selección de preguntas, el análisis de datos y la toma de decisiones sobre éstos Facilita la cooperación del grupo, su unidad y su autorresponsabilidad	Aumenta la credibilidad de los efectos directos de la evaluación a ojos del personal del programa y la probabilidad de que se utilicen los efectos directos Empodera a las partes interesadas y aumenta la capacidad local	Requiere evaluadores con buenas destrezas de facilitación, colaboración y manejo de conflictos, lo mismo que destrezas para capacitar a los participantes en técnicas de evaluación Se considera a menudo como no independiente
Cartografía de efectos directos	Complementa formas más tradicionales de evaluación, centrándose en el cambio en el comportamiento relacionado; el foco está en las personas y los efectos directos y no en el logro de impactos para el desarrollo	Observa el comportamiento resultante de múltiples esfuerzos para mejorar el funcionamiento de los proyectos, programas o políticas, ofreciendo nuevas herramientas, técnicas y recursos Identifica socios de demarcación para evaluar su influencia en el cambio	Sale del "Downstream Outcome Blues" (*ver* texto)	Tiene demanda de evaluación de impacto con independencia y responsabilidad

(Continúa en la página siguiente)

08/11/2010 03:27:48 p.m.

001El camino.indd 196

(Continuación cuadro 5.3)

Enfoque de evaluación	*Propósito / Filosofía*	*Características / Actividades*	*Fortalezas*	*Problemas*
		Describe cambios en la forma como se comportan las personas a través de acciones como individuos o dentro de grupos o de organizaciones		
Evaluación rápida	Suple la necesidad de evaluaciones rápidas y de bajo costo; generalmente busca recopilar sólo la información más esencial para ofrecer indicaciones sobre el desempeño	Utiliza usualmente el enfoque sistemático y semiestructurado Utiliza usualmente el análisis de documentos, la entrevista y cortas visitas al sitio Debe utilizar más de una fuente de información Requiere buenas destrezas de escucha y toma de anotaciones	Sirve de puente entre la recopilación formal e informal de datos: se utiliza mejor cuando examina procesos y problemas	Es usualmente de corto alcance Suministra información limitada y mayormente descriptiva
Síntesis de evaluación y metaevaluación	Localiza y clasifica la calidad de las intervenciones pertinentes y combina o resume sus efectos directos; es útil cuando muchas evaluaciones sobre una intervención ya se han efectuado	Ubica todos los estudios pertinentes Fija criterios para determinar la calidad de los estudios Incluye sólo estudios de calidad Combina efectos directos o de otra manera sintetiza hallazgos de los estudios sobre medidas básicas Puede utilizarse con datos cualitativos o cuantitativos	Utiliza la evaluación y la investigación disponibles, haciéndolas menos costosas Crea una base mayor y más fuerte para evaluar el impacto	Puede tener dificultad para ubicar los estudios pertinentes y obtener datos Lleva algún riesgo de sesgo al seleccionar evaluaciones "de calidad" (los riesgos pueden ser mitigados por el uso de paneles)
Centrada en el uso	Juzga la intervención por su utilidad y cómo se utiliza realmente	Empieza con la identificación y la organización específica y pertinente de quienes toman las decisiones y usuarios de la información (audiencia no vaga y pasiva), que utilizarán la que produce la evaluación	Incorpora el uso de la evaluación, centrándose desde un principio en el uso previsto por parte de los usuarios previstos	Enfatiza en el cliente primario; puede no ser incluyente de otras partes interesadas

(Continúa en la página siguiente)

08/11/2010 03:27:48 p.m.

001El camino.indd 197

(Continuación cuadro 5.3)

Enfoque de evaluación	*Propósito / Filosofía*	*Características / Actividades*	*Fortalezas*	*Problemas*
Para el empoderamiento	Utiliza conceptos, técnicas y hallazgos de evaluación para fomentar el mejoramiento y la autodeterminación	Va más allá de la evaluación participativa, reconociendo un profundo respeto por la capacidad de las personas de crear conocimiento sobre su propia experiencia, y desarrolla soluciones para los problemas que encuentran	Aboga por los grupos desfavorecidos	Cambia el papel del evaluador por el de abogado (puede considerarse la evaluación como política y sesgada)
Realista	Proporciona un marco coherente y consistente para evaluar programas a través de la descripción del contexto; examina cómo debe funcionar un programa y si existe una relación causal entre el programa y el resultado previsto	Derivada de una amplia variedad de enfoques de investigación y evaluación; utiliza parte o todos los demás enfoques Utiliza investigación cualitativa, cuantitativa, o ambas	Esboza la teoría subyacente en el cómo podría funcionar la intervención, en qué condiciones y para quién	Sabe que debe determinarse individualmente el modelo de cada programa
Incluyente	Incluye a las partes interesadas menos favorecidas como parte de la investigación sistemática del mérito o valor de un proyecto, un programa o una política	Genera datos de las partes interesadas menos favorecidas; no incluye a quienes se han incluido tradicionalmente en las evaluaciones	Se centra en las partes interesadas menos favorecidas y las diferencias en la prestación de los servicios	Cambia el papel del evaluador por el de abogado (puede considerarse la evaluación como política y sesgada)
Sobre los beneficiarios	Mejora el impacto de las operaciones para el desarrollo, obteniendo las opiniones de los beneficiarios previstos con respecto a una intervención planeada o en curso	Incluye a los beneficiarios del proyecto, aumentando su sentido de propiedad	Utiliza el enfoque de colaboración que construye propiedad y capacidad, y mejora el valor de la evaluación, capturando las opiniones de los beneficiarios	Separa la consideración de los problemas ambientales y sus consecuencias sociales
Horizontal	Combina procesos internos de evaluación con revisión externa de pares para neutralizar relaciones de poder desequilibradas que predominan en las evaluaciones externas tradicionales	Es utilizada para el aprendizaje y mejoramiento de la evaluación y de metodologías para el desarrollo que se encuentran en desarrollo	Combina las fortalezas de la autoevaluación y la revisión de pares externos	Entiende que las revisiones de pares pueden ser mutuamente positivas

Fuente: los autores.

08/11/2010 03:27:48 p.m.

Actividad del capítulo 5

Ejercicio de aplicación 5.1: escoger el enfoque de evaluación más apropiado

Seleccione un enfoque de evaluación para analizar cada una de las asignaciones descritas a continuación. Describa el enfoque, liste los beneficios que proporciona y los problemas que presenta y explique sus razones para escogerlo.

1. Evalúe el foco estratégico de asistencia técnica a un país, con base en los hallazgos de cinco estudios de países efectuados por distintas organizaciones para el desarrollo.
2. Identifique intervenciones educativas exitosas implementadas en los proyectos y programas de su organización para mejorar los sistemas educativos en la región.
3. Evalúe los problemas más significativos concernientes a los recursos naturales y el sector ambiental de un país.
4. Evalúe el desarrollo del sector arrocero de un país, incluyendo la importancia del arroz en el contexto cultural, social y económico; los sistemas de producción, las restricciones que tienen los agricultores, la investigación realizada, las tecnologías desarrolladas y las prioridades futuras para el desarrollo adicional del arroz.
5. Analice las evaluaciones realizadas por una organización para el desarrollo que ha recibido millones de dólares en financiación para investigación agrícola internacional en los últimos 30 años.

Referencias y lecturas adicionales

Alkin, Marvin y Christina Christie. 2004. "An Evaluation Theory Tree". En *Evaluation Roots: Tracing Theorist Views and Influences*, ed. M. Alkin. 12-65, Thousand Oaks, CA: Sage Publications.

Banco Mundial. 1996. *World Bank Participation Sourcebook.* Washington, DC. http://www.worldbank.org/wbi/sourcebook/sba108.htm#D.

———. 2004. "Social Assessment". En *Turning Bureaucrats into Warriors.* 135-38. Washington, DC. http://www.worldbank.org/afr/aids/gom/manual/GOM-Chapter percent2024.pdf.

Chambers, R. 1991. "Shortcut and Participatory Methods for Gaining Social Information for Projects". En *Putting People First: Sociological Variables in Rural Development,* 2ª ed., ed. M. M. Cernea, 515-37. Washington, DC: Banco Mundial.

Chelimsky, E. y L. G. Morra. 1984. "Evaluation Synthesis for the Legislative User". En *Issues in Data Synthesis,* ed. W. H. Yeaton y P. M. Wortman, 75-89. New Directions for Program Evaluation N° 24. San Francisco: Jossey-Bass.

Christie, Christina y Marvin Alkin. 2004. "Objectives-Based Evaluation". En *Encyclopedia of Evaluation*, ed. Sandra Mathison. 281-85. Thousand Oaks, CA: Sage Publications.

CIDA (Canadian International Development Agency). 2004. *CIDA Evaluation Guide 2004.* Ottawa.

Cousins, J. B. y L. M. Earl, eds. 1995. *Participatory Evaluation in Education.* Bristol, PA: Falmer Press.

Dalal-Clayton, Barry y Stephen Ba. 2002. *Sustainable Development Strategies: A Resource Book*. Sterling, VA: Earthscan Publications. http://www.nssd.net/res_book.html#contents.

DFID (Department for International Development). 2000. *Environment: Mainstreamed or Sidelined?* Environmental Evaluation Synthesis Study EV626, enero, Londres. http://www.dfid.gov.uk/aboutdfid/performance/files/ev626s.pdf.

Duignan, Paul. 2007. *Introduction to Strategic Evaluation: Section on Evaluation Approaches, Purposes, Methods y Designs*. http://www.strategicevaluation.info/se/documents/104f.html.

Earl, Sarah, Fred Carden y Terry Smutylo. 2001. *Outcome Mapping: Building Learning and Reflection into Development Programs.* International Development Research Centre, Ottawa. http://www.dgroups.org/groups/pelican/docs/Mapping_M&E_capacity_080606.pdf.

ECDPM (European Centre for Development Policy Management). 2006. *Study on Capacity, Change and Performance: Mapping of Approaches Towards M&E of Capacity and Capacity Development.* Maastricht.

Eerikainen, Jouni y Roland Michelevitsh. 2005. "Environmental and Social Sustainability. Methodology and Toolkit: Various Approaches". International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio.

FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). 1997. "Rapid Rural Appraisal". En *Marketing Research and Information Systems Marketing and Agribusiness.* Roma. http://www.fao.org/docrep/W3241E/w3241e09.htm.

Fetterman, David M. 2001. *Foundations of Empowerment Evaluation.* Thousand Oaks, CA: Sage Publications.

Fetterman, David M., S. Kaftarian y A. Wandersman, eds. 1996. *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability.* Thousand Oaks, CA: Sage Publications.

Fetterman, David M. y Abraham Wandersman. 2004. *Empowerment Evaluation Principles in Practice.* Nueva York: Guilford Publications.

———. 2007. "Empowerment Evaluation: Yesterday, Today y Tomorrow". *American Journal of Evaluation* 28 (2):179-98. http://homepage.mac.com/profdavidf/documents/EEyesterday.pdf.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* Nueva York: Pearson.

Gariba, Sulley. 1998. "Participatory Impact Assessment: Lessons From Poverty Alleviation Projects in Africa". En *Knowledge Shared: Participatory Evaluation in Development Cooperation*, ed. Edward T. Jackson y Yussuf Kassam, 64-81. Bloomfield, CT: Kumarian Press.

Gill, M. y A. Spriggs. 2002. *The Development of Realistic Evaluation Theory through the Evaluation of National Crime Prevention Programmes.* http://www.evaluationcanada/distribution/20021010_gill_martin_spriggs_angela. pdf.

Glass, Gene V. y Mary Lee Smith. 1979. "Meta-Analysis of Research on Class Size and Achievement". *Educational Evaluation and Policy Analysis* 1 (1): 2-16.

Hirschheim, R. y S. Smithson. 1988. "A Critical Analysis of Information Systems Evaluation". En *IS Assessment: Issues and Changes*, eds. N. Bjorn-Andersen y G. B. Davis. Amsterdam: North-Holland.

ISO (International Organization for Standardization). 1999. *Environmental Management: Performance Evaluation Guidelines.* ISO 14301, Ginebra.

Johnston, Timothy y Susan Stout. 1999. *Investing in Health: Development in Health, Nutrition y Population Sector.* Banco Mundial, Operations Evaluation Department, Washington, DC. http://wbln0018.worldbank.org/oed/oeddoclib.nsf/6e14e487e87320f785256808006a001a/daf8d4188308862f852568420062f332/US$FILE/HNP.pdf.

Khon Kaen University. 1987. *Proceedings of the 1985 International Conference on Rapid Rural Appraisal.* Rural Systems Research and Farming Systems Research Projects, Khon Kaen, Tailandia.

Kretzmann, John P., John L. McKnight y Deborah Puntenney. 1996. A Guide to Massing Local Business Assets and Modulizing Local Business Capabilities. Skokie, IL: ACTA Publications.

Kudat, Ayse y Bykebt Ozbilgin. 1999. *Azerbaijan Agricultural Development and Credit Program.* http://lnweb18.worldbank.org/ESSD/sdvext.nsf/61ByDocName/AzerbaijanAgriculturalDevelopmentandCreditProject/US$FILE/AzerbaijanAgriculturalDevelopmentandCreditProject-424KbPDF.pdf.

Kumar, Krishna, ed. 1993. *Rapid Appraisal Methods.* Washington, DC: Banco Mundial.

Light, R. J. y D. B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research.* Cambridge, MA: Harvard University Press.

Mertens, D. 1999. "Inclusive Evaluation: Implications of Transformative Theory for Evaluation". *American Journal of Evaluation* 20 (1): 1-14.

Narayan, Deepa. 1996. *Toward Participatory Research.* Documento técnico del Banco Mundial 307, Banco Mundial, Washington, DC. http:// www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1996/04/01/000009265_3980625172923/Rendered/PDF/multi0page.pdf.

Nay, J. y P. Kay. 1982. *Government Oversight and Evaluability Assessment.* Lexington, MA: Heath.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 1997. *Searching for Impact and Methods: NGO Evaluation Synthesis Study.* Development Assistance Committee, París. http://www.eldis.org/static/DOC5421.htm.

Patton, Michael Q. 1990. *Qualitative Evaluation and Research Methods.* 2ª ed. Thousand Oaks, CA: Sage Publications.

———. 1997. *Utilization-Focused Evaluation: The New Century Text.* Thousand Oaks: CA: Sage Publications.

———. 1997. "Toward Distinguishing Empowerment Evaluation and Placing It in a Larger Context". *Evaluation Practice* 18 (2):147-63.

———. 2002. *Qualitative Evaluation and Research Methods*, 3ª ed. Thousand Oaks, CA: Sage Publications.

———. 2008. *Utilization-Focused Evaluation*. 4ª ed. Thousand Oaks, CA: Sage Publications.

Paulmer, Hubert E. 2005. "Evaluation Guidelines of International Development Aid Agencies: A Comparative Study". International Rural Planning and Development, School of Environmental Design and Rural Development, University of Guelph, Ontario.

Pawson, Ray y Nick Tilley. 2004. "Realistic Evaluation". En *Encyclopaedia of Evaluation*, ed. Sandra Matthieson, 359-67. Thousand Oaks, CA: Sage Publications.

Picciotto, Robert. 2007. "The New Environment for Development Evaluation". *American Journal of Evaluation* 28: 509-21.

Preskill, Hallie y Darlene Russ-Eft. 2005. *Building Evaluation Capacity: 72 Activities for Teaching and Training.* Thousand Oaks, CA: Sage Publications.

Salmen, Lawrence F. 1999. *Beneficiary Assessment Manual for Social Funds.* Banco Mundial, Social Protection Team Human Development Network, Washington, DC. http://lnweb18.worldbank.org/ESSD/sdvext.nsf/07ByDocName/BeneficiaryAssessmentManualforSocialFunds/US$FILE/percent5BEnglish percent5D+Beneficiary+Assessment+Manual.pdf.

Sanders, J. R. 1997. "Cluster Evaluation". En *Evaluation for the 21st Century: A Handbook*, eds. E. Chelimsky y W. R. Shadish Jr., 396-401. Thousand Oaks, CA: Sage Publications.

Scrimshaw, N. y G. R. Gleason. 1992. *Rapid Assessment Procedures: Qualitative Methodologies for Planning and Evaluation of Health Related Programs.* Boston: International Nutrition Foundation for Developing Countries.

Scriven, Michael. 1972a. "Objectivity and Subjectivity in Educational Research". En *Philosophical Redirection Educational Research: The Seventy-First Yearbook of the National Society for the Study of Education*, ed. L. G. Thomas. Chicago: University of Chicago Press.

001El camino.indd 200 08/11/2010 03:27:48 p.m.

———. 1972b. "Pros and Cons about Goal-Free Evaluation". *Evaluation Comment* 3: 1-7.
———. 1991. *Evaluation Thesaurus.* 4ª ed. Thousand Oaks, CA: Sage Publications.
Smith, M. F. 1989. *Evaluability Assessment: A Practical Approach.* Boston: Kluwer Academic Press.
Smith, Mary Lee y Gene V. Glass. 1980. "Meta-Analysis of Research on Class Size and Its Relationship to Attitudes and Instruction". *American Educational Research Journal* 17: 419-33.
Smith, Nick L. y Paul R. Brandon, eds. 2007. *Fundamental Issues in Evaluation.* Nueva York: Guilford Press.
Smutylo, Terry. "The Output Outcome Downstream Impact Blues". http://www.idrc.ca/en/ev-65284-201-1-DO_TOPIC.html.
Thiele, G., A. Devaux, C. Velasco y D. Horton. 2007. "Horizontal Evaluation: Fostering Knowledge Sharing and Program Improvement within a Network". *American Journal of Evaluation* 28 (4): 493-508.
Thiele, G., A. Devaux, C. Velasco y K. Manrique. 2006. "Horizontal Evaluation: Stimulating Social Learning among Peers". International Potato Center, Papa Andina Program, Lima, Perú. Borrador, mayo 18. http://www.dgroups.org/groups/pelican/docs/Hor_Evaln_18_05.doc?ois=no.
Tilley, Nick. 2000. *Realistic Evaluation: An Overview.* Documento presentado en la Founding Conference of the Danish Evaluation Society, Nottingham Trent University, septiembre. http://www.danskevalueringsselskab.dk/pdf/Nick percent20Tilley.pdf.
Turpin, R. S. y J. M. Sinacore, eds. 1991. *Multisite Evaluations.* New Directions for Program Evaluation N° 50. San Francisco, CA: Jossey-Bass.
U.S. GAO (General Accounting Office). 1990. *Prospective Evaluation Methods: The Prospective Evaluation Synthesis.* Washington, DC. http://www.gao.gov/special.pubs/10_1_10.PDF.
———. 1992. *The Evaluation Synthesis.* Washington, DC. http://www.gao.gov/special.pubs/pemd1012.pdf.
Wandersman, A., J. Snell-Johns, B. E. Lentz, D. M. Fetterman, D. C. Keener, M. Livet, P. S. Imm y P. Flaspoler. 2005. "The Principles of Empowerment Evaluation". En *Empowerment Evaluation Principles in Practice,* eds. D. M. Fetterman y A. Wandersman, 27-41. Nueva York: Guilford.
Wholey, J. S. 1987. "Evaluability Assessment: Developing Program Theory". En *Using Program Theory in Evaluation*, ed. L. Bickman, 77-92. New Directions for Program Evaluation N° 33. San Francisco: Jossey-Bass.
Wholey, J. S., H. P. Hatry y K. E. Newcomer, eds. 1994. *Handbook of Practical Program Evaluation.* San Francisco: Jossey-Bass.

Sitios web

Banco Mundial. 2009. *Social Analysis.* http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTSOCIALANALYSIS/0,,menuPK:281319~pagePK:149018~piPK:149093~theSitePK:281314,00.html.
CIF (Corporación Financiera Internacional). *Environmental and Social Policies and Guidelines.* http://www.ifc.org/ifcext/enviro.nsf/Content/PoliciesandGuidelines.
Equator Principles. http://www.equator-principles.com/, http://www.ifc.org/ifcext/equatorprinciples.nsf/Content/ThePrinciples.
Frechtling, Joy y Laure Sharp Westat, eds. 1997. *User-Friendly Handbook for Mixed-Method Evaluations.* Washington, DC: National Science Foundation. http://www.nsf.gov/pubs/1997/nsf97153/start.htm.
ISO (International Organization for Standardization). *ISO 14031.* http://www.iso.org/iso/catalogue_detail?csnumber=23149, http://www.iso-14001.org.uk/iso-14031.htm y http://www.altech-group.com/ftp/EPEarticle.pdf.

IUCN (World Conservation Union). *Sustainability Assessment.* http://www.iucn.org/themes/eval/search/iucn/sustassess.htm.

Evaluación grupal

AusAid. 2004. *Governance in PNG: A Cluster Evaluation of Three Public Sector Reform Activities.* Evaluation and Review Series 35, Sydney. http://www.ausaid.gov.au/publications/pdf/governance_in_png_qc35.pdf.

Evaluación para el empoderamiento

Empowerment Evaluation Blog. http://www.eevaluation.blogspot.com/.

Evaluación ambiental y social

Lanco Amarkantak Thermal Power. *Environmental and Social Review.* Informe preparado para la IFC Board of Directors. http://www.ifc.org/ifcext/spiwebsite1.nsf/1ca07340e47a35cd85256efb00700ceeC126975A64D3306E852572A0004807BD.

Evaluación de la evaluabilidad

Lewin Group. 1997. *An Evaluability Assessment of Responsible Fatherhood Programs.* http://fatherhood.hhs.gov/evaluaby/intro.htm.

Síntesis de evaluación

Kruse, S. E., T. Kyllönen, S. Ojanperä, R. C. Riddell y J. Vielaj. *Searching for Impact and Methods: NGO Evaluation Synthesis Study.* Institute of Development Studies, University of Helsinki. http://www.eldis.org/static/DOC5421.htm.

Evaluación basada en objetivos

IFAD (International Fund for Agricultural Development). *Country Program Evaluation of the People's Republic of Bangladesh.* http://www.ifad.org/evaluation/public_html/eksyst/doc/country/pi/bangladesh/bangladesh.htm.

Evaluación sin tener en cuenta las metas

Evaluation Center, Western Michigan University. http://www.wmich.edu/evalctr/project-pub.html.

Metaevaluación

Campbell Collaboration Online Library. http://www.campbellcollaboration.org/frontend.aspx.

Evaluación en sitios múltiples

Australian HIV Test Evaluation Group. 1995. *Multisite Evaluation of Four Anti-HIV-1/HIV-2 Enzyme Immunoassays.* http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=7882108&dopt=Abstract/.

SRI International. 2001. *A Multisite Evaluation of the Parents as Teachers (PAT) Project.* http://policyweb.sri.com/cehs/projects/displayProject.jsp?Nick=pat.

Cartografía de resultados

African Health Research Fellowship Program. *Strengthening Leadership Capacity to Improve the Production and Use of Health Knowledge in Africa.* http://www.idrc.ca/fr/ev-34425-201-1-DO_TOPIC.html.

Evaluación participativa

Contra Costa Health Services. 2009. *Preventing Chronic Disease: A Participatory Evaluation Approach.* http://www.cchealth.org/groups/chronic_disease/guide/evaluation.php.

Community Integrated Pest Management. 2001. *Picturing Impact: Participatory Evaluation of Community IPM in Three West Java Villages.* http://www.communityipm.org/docs/Picturing percent20Impact/Picturing percent20Impact percent20top percent20page.html.

Evaluación prospectiva

Glewwe, Paul, Michael Kremer y Sylvie Moulin. 1998. *Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya.* http://www.econ.yale.edu/~egcenter/infoconf/kremer_paper.pdf.

Evaluación rápida

UNICRI (United Nations Interregional Crime and Justice Research Institute), and ACI (Australian Institute of Criminology). *Global Programme against Trafficking in Human Beings, Rapid Assessment: Human Smuggling and Trafficking from the Philippines.* http://www.unodc.org/pdf/crime/trafficking/RA_UNICRI.pdf.

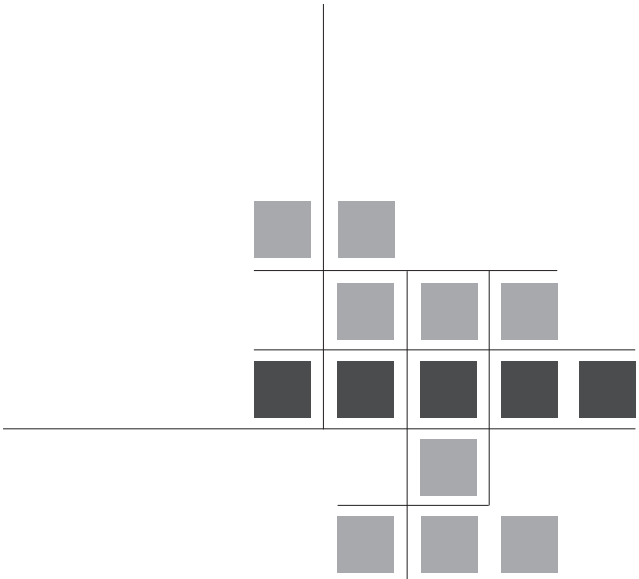
Evaluación social

Banco Mundial. 1996. "Morocco: Fez Medina Rehabilitation Project". En *World Bank Participation Sourcebook.* Washington, DC: Banco Mundial. http://www.worldbank.org/wbi/sourcebook/sba108.htm#D.

Evaluación centrada en el uso

Evaluation Center, Western Michigan University. *Utilization-Focused Evaluation Checklist.* http://www.wmich.edu/evalctr/checklists/ufechecklist.htm.

001El camino.indd 204 08/11/2010 03:27:49 p.m.



DISEÑO Y REALIZACIÓN

"No hay límite para lo complicado que pueden ponerse las cosas, pues una cosa siempre conduce a otra".

—E.B. WHITE

Capítulo 6. Desarrollo de preguntas de la evaluación e inicio de la matriz de diseño

- Fuentes de preguntas
- Tipos de preguntas
- Identificación y selección de preguntas
- Desarrollo de buenas preguntas
- Diseño de la evaluación

Capítulo 7. Selección de diseño para preguntas de evaluación de causa y efecto, normativas y descriptivas

- Conectar las preguntas al diseño
- Diseño de preguntas de causa y efecto
- Diseño de preguntas descriptivas
- Diseño de preguntas normativas
- Necesidad de diseño más riguroso

001El camino.indd 205 08/11/2010 03:27:49 p.m.

Capítulo 8. Selección y construcción de instrumentos de recopilación de datos

- Estrategias de recopilación de datos
- Características de las buenas medidas
- Datos cuantitativos y cualitativos
- Herramientas para la recopilación de datos

Capítulo 9. Selección de la estrategia de muestreo

- Introducción al muestreo
- Tipos de muestras: aleatorias y no aleatorias
- Establecer el tamaño de la muestra
-

Capítulo 10. Planeación y realización de análisis de datos

- Estrategia de análisis de datos
- Análisis de datos cualitativos
- Análisis de datos cuantitativos
- Vínculo de datos cuantitativos y cualitativos

CAPÍTULO 6

Desarrollo de preguntas de la evaluación e inicio de la matriz de diseño

De cinco capítulos este es el primero en donde se tratan pasos específicos sobre el diseño de una evaluación. En él se discuten los tipos de preguntas de la evaluación, y se explica cuándo utilizar cada uno. En el capítulo también se cubre cómo escribir y estructurar preguntas buenas.

Fuentes de preguntas

¿Por qué es tan importante seleccionar las preguntas para la evaluación? Una de las razones es que las preguntas le dan dirección a la evaluación y al diseño seleccionado de ésta (se tratan estas implicaciones en forma más completa en el capítulo 7). Las preguntas de la evaluación constituyen un elemento crítico que contribuye a mejorar los esfuerzos de individuos y grupos clave, la toma de decisiones y la información al público. La reflexión y la investigación cuidadosas son necesarias para completar el proceso crítico de identificar y definir las preguntas de una evaluación (Fitzpatrick, Sanders y Worthen, 2004).

En la evaluación los evaluadores hacen preguntas para conocer el proyecto, el programa o la política que se evalúa. Al desarrollar las preguntas, un problema frecuente que se

001El camino.indd 207 08/11/2010 03:27:49 p.m.

encuentra es el de asumir que todos los implicados comparten el mismo conocimiento de las metas de la intervención. Si, por ejemplo, la pregunta es "¿Ayudó el programa a los participantes?", distintas partes interesadas pueden interpretar las palabras *ayudar* y *participantes* en forma diferente. Lograr un acuerdo sobre la teoría del cambio tratada en el capítulo anterior puede remediar este problema.

Fitzpatrick, Sanders y Worthen (2004) desarrollaron una lista de las fuentes que los evaluadores deberían utilizar para asegurarse de obtener puntos de vista diversos:

- Preguntas, preocupaciones y valores de los partes interesadas.
- Modelos de evaluación.
- Marcos y enfoques, incluso enfoques heurísticos (de prueba y error).
- Hallazgos de investigación y evaluación y temas importantes encontrados en la literatura.
- Estándares profesionales, listas de verificación, lineamientos, instrumentos o criterios desarrollados o utilizados en otra parte.
- Opiniones y conocimiento de consultores expertos.
- El propio juicio profesional del evaluador.

En el capítulo 4 se cubrieron la identificación y el funcionamiento con las partes interesadas para solicitar sus opiniones sobre asuntos que consideran importantes para su evaluación, y también se cubrió la importancia de revisar estudios anteriores de investigación y evaluación para identificar las preguntas. Se enfatizó en el desarrollo y el uso de modelos de la teoría del cambio a fin de identificar áreas de enfoque para la evaluación. Las preguntas provienen de los mayores supuestos subyacentes en el modelo.

En el gráfico 6.1 se presentan los tipos de preguntas de evaluación que deben hacerse en los diferentes puntos de una cadena causal. Las preguntas genéricas que aparecen en la parte inferior del diagrama muestran que pueden obtenerse preguntas formativas de actividades y productos, y las preguntas finales de los efectos directos intermedios y de largo plazo. Las preguntas derivadas de los resultados a corto plazo pueden considerarse como formativas o finales. Las preguntas deberían fluir de los mayores supuestos del modelo de lógica sobre cómo funcionará el programa, y qué beneficios o efectos directos se espera lograr. Las preguntas provienen también de la revisión de investigación de evaluaciones concluidas de programas similares, al igual que de las diversas perspectivas de las partes interesadas sobre el proyecto, el programa o la política, según se trató en el capítulo 4.

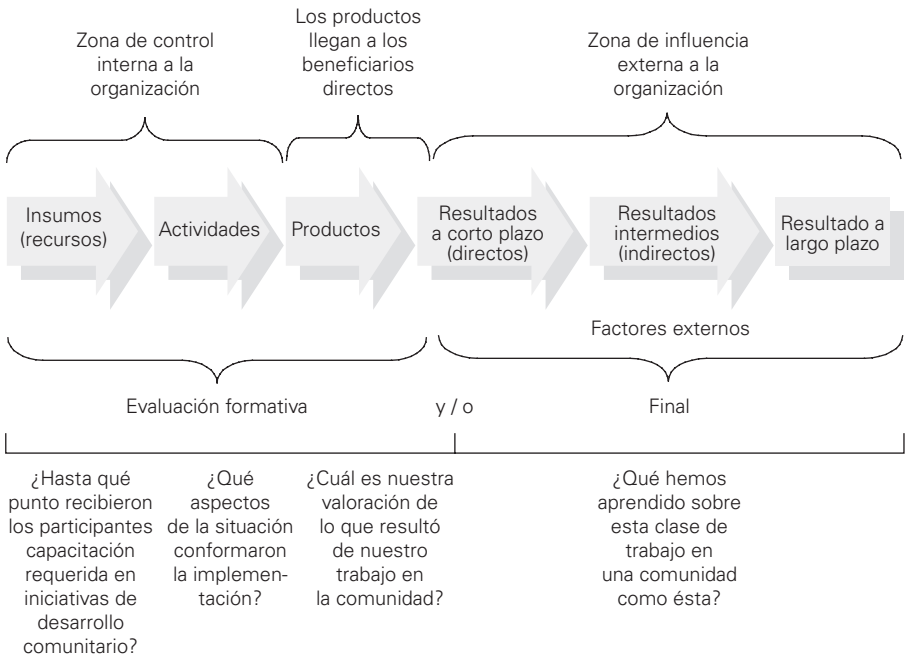
Tipos de preguntas

Al planear una evaluación, pueden considerarse muchas preguntas, aunque todas deben definirse claramente en forma mesurable.

Las preguntas pueden agruparse en tres categorías: descriptivas, normativas y de causa y efecto. Los tipos de preguntas –junto con los datos, el tiempo y el dinero disponible para realizar la evaluación– orientarán la clase de diseño seleccionado.

001El camino.indd 208 08/11/2010 03:27:49 p.m.

Gráfico 6.1 *Uso de un modelo lógico para enmarcar las preguntas de la evaluación*



Preguntas descriptivas

Pregunta descriptiva: Pregunta con la que se busca determinar lo que es.

Con las preguntas descriptivas se busca determinar lo que es y pueden describir aspectos de un proceso, una condición, un conjunto de puntos de vista, o un conjunto de relaciones o de redes organizativas. Patton (2002) se refiere a las preguntas descriptivas como el fundamento de las evaluaciones.

Las preguntas descriptivas:

- Buscan comprender o describir un programa o proceso.
- Proporcionan una "instantánea" de lo que es.
- Son directas (quién, qué, dónde, cuándo, cómo y cuántos).
- Pueden utilizarse para describir insumos, actividades y productos.
- Se utilizan con frecuencia para recopilar la opinión de los clientes del programa.

Algunos ejemplos de preguntas descriptivas son:

- ¿Cuáles son las metas del programa desde la perspectiva de las diferentes partes interesadas?

- ¿Cuáles son las actividades primarias del programa?
- ¿Cómo ingresan las personas al programa?
- ¿Dónde se ha implementado el programa?
- ¿Qué servicios presta el programa para hombres? ¿Para mujeres?
- ¿Qué efectos tiene el programa en los participantes?
- ¿Hasta qué punto el diseño del programa refleja enseñanzas aprendidas en programas similares anteriores?
- ¿Hasta qué punto existen diferencias entre los sitios del programa en cuanto a la implementación del programa?
- ¿Qué calificación tienen los proveedores del servicio?
- ¿Cuándo se implementó el programa?
- ¿Cuántas mujeres participaron en el programa?
- ¿Cómo se compara el costo del programa con los costos de programas similares?
- ¿Cuáles son los canales de comunicación informal dentro de la organización?
- ¿Qué tan útil fue el programa para los participantes?

Las preguntas de evaluación sobre formulación de políticas son, con mucha frecuencia, preguntas descriptivas. Rist (1994) identifica un proceso de "ciclo de política" de tres fases, durante las cuales se hacen las preguntas (cuadro 6.1).

Preguntas normativas

Con las *preguntas normativas* se compara lo que es con lo que debería ser. Se compara la situación actual frente a un objetivo, una meta o un estándar de comparación. Estas preguntas son similares en la orientación del cumplimiento a las que se hacen con mucha frecuencia al efectuarse la auditoría. Son las siguientes:

- ¿Estamos haciendo lo que se supone hagamos?
- ¿Estamos logrando nuestro objetivo?
- ¿Logramos lo que dijimos que lograríamos?

Si el programa tiene un sistema de seguimiento basado en resultados, con indicadores y objetivos, y tiempo fijado para lograrlos, las preguntas normativas pueden utilizarse para responder preguntas sobre insumos, actividades y productos.

En ocasiones un programa tiene objetivos, pero no tiene criterios aparentes para determinar cómo se medirán o lograrán. No se han establecido indicadores u objetivos. En tal caso, el evaluador tiene varias opciones, algunas mejores que otras (recuadro 6.1).

La búsqueda de estándares o de criterios empieza generalmente con los criterios que se encuentran en los documentos que autorizan el programa; por ejemplo, la legislación o los documentos de aprobación de la junta directiva. Los criterios pueden especificarse también como indicadores con objetivos específicos en los sistemas de gestión basada en

Preguntas normativas: Son las que comparan lo que es con lo que debería ser.

Cuadro 6.1 *Preguntas cualitativas para plantear durante diferentes fases del ciclo de política*

Fase	*Ejemplos*
Formulación de política	• ¿Qué ha ocurrido anteriormente como respuesta a esta condición o problema? • ¿Qué se conoce sobre esfuerzos anteriores que ayudaría a quien toma las decisiones a elegir entre las opciones actuales?
Implementación de política	• ¿Qué información sobre el problema o la condición indujo la respuesta de la política o el programa? • ¿Qué esfuerzos hizo la organización o la institución para responder a la iniciativa? • ¿Qué calificaciones tienen los responsables para la implementación? • ¿Hasta qué punto han mostrado interés la administración y el personal? • ¿Qué controles existen sobre la asignación de los recursos? • ¿La estructura organizativa refleja en forma adecuada las demandas sobre la organización para responder a esta iniciativa? • ¿Qué medios existen en la organización para decidir sobre demandas que compiten entre sí? • ¿Qué clase de sistemas de retroalimentación existen para asistir a los administradores?
Rendición de cuentas de la política	• ¿Qué pertinencia tiene el programa o la política en cuanto a las circunstancias actuales? • ¿Qué tan fuerte fue el interés en la rendición de cuentas?

Fuente: adaptado de Rist, 1994.

resultados. Otras fuentes que pueden establecer los estándares son: los sistemas de acreditación, los paneles de alto nivel, las organizaciones profesionales y otras comisiones.

Algunos ejemplos de preguntas normativas son:

- ¿Gastamos lo presupuestado?
- ¿Logramos la meta de admitir 5.000 estudiantes por año?
- ¿Vacunamos al 80% de los niños según se requería?
- ¿Cumplimos el objetivo de drenar 10.000 hectáreas de terreno?
- ¿Fue justo y equitativo el proceso de selección de participantes?

Preguntas de causa y efecto

Pregunta de causa y efecto: Pregunta que determina la diferencia que representa una intervención.

Las preguntas de causa y efecto determinan la diferencia que representa una intervención. Se conocen a menudo como *preguntas de resultado, impacto* o *atributivas*, y con ellas se intenta medir los cambios debidos a la intervención. Con las preguntas de causa y efecto se busca establecer los efectos de un proyecto, un programa o una política. Son preguntas del tipo "¿entonces qué?" y con ellas se pregunta si se lograron los resultados deseados como resultado del programa.

Recuadro 6.1 *Evaluar una intervención que no tiene estándares*

Considérese un programa multisectorial cuyos objetivos incluyen los siguientes:

- Mejorar los puntajes de lectura de los niños en distritos escolares seleccionados.
- Aumentar la percepción del riesgo de VIH/Sida y los métodos para prevenirlo en una región.
- Aumentar el número de microempresas en una aldea y sus beneficios.

Al evaluador le gustaría saber exactamente qué significan estos objetivos (qué proporción de niños mejorará su puntaje en lectura, en quiénes aumentará la percepción de VIH/Sida y cómo lo sabremos, cuántas microempresas más se formarán y en qué medida), pero el programa no ha definido sus objetivos en estos términos. ¿Qué puede hacer el evaluador en estas circunstancias?

Un enfoque es el de trabajar con los "propietarios" del programa, o sea, los funcionarios responsables administrativamente del programa o de su implementación. Estos funcionarios deberían poder indicar un nivel razonable del desempeño que debería obtener el programa.

Una preocupación con este enfoque es que un grupo puede no aceptar los estándares que ha fijado otro grupo. El personal con responsabilidad de supervisión, por ejemplo, puede no estar de acuerdo con el estándar propuesto por los ejecutores del programa, y podría argumentar que los estándares fijados son demasiado bajos.

Otro enfoque es traer uno o más expertos para que se pongan de acuerdo sobre un estándar que pueda utilizarse. Con este enfoque un problema potencial es que el estándar puede reflejar los sesgos personales del experto. Esta crítica podría disminuir si se utilizan varios expertos. En tal caso, es importante que el grupo de expertos se considere como políticamente neutral o equilibrado, y que los expertos no tengan participación anterior en el programa.

La alternativa más débil y arriesgada es que la evaluación fije el estándar. Este enfoque debe evitarse porque presenta dificultades para el evaluador. Quienes están dentro del programa podrían argumentar que los estándares son demasiado altos o demasiado bajos, o que dichos estándares no pueden fijarse después del hecho y utilizarse para juzgar su desempeño.

Fuente: los autores.

Los modelos de la teoría programática del cambio describen los efectos directos deseados y los impactos de una intervención, pero los efectos directos pueden o no expresarse como preguntas de causa y efecto. Por ejemplo, en un programa para presentar a los agricultores una semilla nueva y mejorada, una pregunta de resultado podría ser la de si aumentó la producción del grano. Según se enunció, ésta podría ser una pregunta descriptiva ya que simplemente se pregunta cuánto aumentó la cosecha. Si en la evaluación se pregunta si la cosecha aumentó como resultado del programa –y no, por ejemplo, como resultado de un clima o un tiempo excepcionalmente ideal para la cosecha del grano–, entonces se trata claramente de una pregunta de causa y efecto. Las preguntas de causa y efecto implican una comparación del desempeño en una o más medidas o unos indicadores, no sólo antes y después de la intervención, sino también con y sin ella.

001El camino.indd 212 08/11/2010 03:27:49 p.m.

Algunos ejemplos de preguntas de causa y efecto son:

- ¿Fue efectiva la estrategia de asociación de los tres países para preservar la biodiversidad de la zona afectada, y se mantuvieron los medios de vida?
- Como resultado del programa de capacitación, ¿los participantes tienen empleos de mejor salario de lo que habría sido de otra forma?
- ¿Redujo la pobreza el programa de microempresas en los municipios en donde operó?
- ¿Con el aumento del impuesto a la gasolina mejoró la calidad del aire?
- ¿El aumento del gobierno a las sanciones financieras para las empresas que violaron la ley redujo el uso de niños de corta edad en la industria de vestuario?
- ¿Qué otros impactos o efectos laterales (positivos o negativos) tuvo esta intervención en la comunidad en general?

Los evaluadores deben plantear tales preguntas en términos de causa y efecto, pues como muchas actividades ocurren al mismo tiempo, es difícil demostrar que los efectos directos se deben sólo, o al menos primordialmente, a la intervención. Al realizar diseños para responder preguntas de causa y efecto, los evaluadores deben tener gran cuidado para eliminar otras explicaciones posibles de los cambios que miden.

En el capítulo 7 se tratan los diseños que pueden utilizarse para responder las preguntas de causa y efecto, y se examinan las clases de análisis necesario para atribuir una relación causal. Dado que es más difícil responder preguntas de causa y efecto que descriptivas o normativas, es importante estar bastante seguro de que se desea y necesita esta forma de pregunta.

En muchas evaluaciones sólo se utilizan preguntas descriptivas y normativas, en particular si se trata de evaluaciones formativas que se centran en la implementación de una intervención. En las evaluaciones que se centran en el impacto se hacen preguntas de causa y efecto, pero usualmente se incluyen también algunas preguntas descriptivas y normativas. En el recuadro 6.2 se ilustra cómo se pueden incluir en una evaluación distintos tipos de preguntas.

Relación entre los tipos de preguntas y las teorías del cambio

¿Cómo se ajustan los tipos de preguntas a las teorías del cambio tratadas en el capítulo anterior? Las preguntas sobre si se han conseguido los insumos de un programa y se han logrado los objetivos en cuanto a sus productos son usualmente normativas: ¿Con los fondos empleados se adquirieron la cantidad requerida de bienes y la provisión de servicios en el tiempo especificado? Las preguntas sobre el logro de los efectos directos trazados como objetivo son también normativas. Preguntar hasta qué punto aumentaron los puntajes en lectura en un período de tiempo es una pregunta descriptiva; preguntar si los puntajes en lectura aumentaron hasta el objetivo propuesto por los administradores del programa es una pregunta normativa. Las preguntas con las cuales se busca probar

Recuadro 6.2 Evaluar políticas e intervenciones utilizando preguntas y respuestas

Mejorar la atención en salud preventiva

Política: Asegurar que todos los niños reciban atención de salud preventiva.
Meta: Reducir la mortalidad infantil y preescolar.

Preguntas de evaluación:

1. ¿Qué porcentaje de niños recibió atención de salud preventiva desde el inicio del programa? (Pregunta descriptiva)
2. ¿Han recibido los grupos previstos de niños de bajos ingresos atención de salud preventiva? (Pregunta normativa)
3. ¿Disminuyeron las tasas de mortalidad infantil como resultado del programa? (Pregunta de causa y efecto).

Capacitar estudiantes de secundaria para el mercado laboral

Política: Asegurar que en los colegios de secundaria se enseñen el conocimiento y las destrezas requeridas para el empleo en el mercado local.
Meta: Asegurar que los graduados puedan obtener trabajo calificado de buen salario.

Preguntas de evaluación:

1. ¿Cómo están preparando a los estudiantes los colegios de secundaria para el trabajo en el mercado local? (Pregunta descriptiva)
2. Después de un año, ¿reciben los graduados salarios mayores que quienes abandonaron el programa? (Pregunta descriptiva)
3. ¿Hasta qué punto están los colegios de secundaria tomando decisiones basados en el mercado en los campos de capacitación requeridos? (Pregunta normativa)
4. ¿Cuánto más ganan los graduados de lo que habrían ganado sin el programa? (Pregunta de causa y efecto)

Vacunación gratis contra el sarampión

Intervención: Las clínicas familiares ofrecen vacunación gratis contra el sarampión a todos los menores de cinco años en tres regiones del país en un año.

Preguntas de evaluación:

1. ¿Cómo llegaron las clínicas a los padres para informarles sobre las vacunaciones gratis para los niños? (Pregunta descriptiva)
2. ¿El programa cumplió el objetivo de proporcionar vacunación gratis contra el sarampión al 100% de los niños menores de cinco años en las tres regiones el año anterior? (Pregunta normativa)
3. ¿El programa utilizó métodos innovadores para llegar a los niños con mayor riesgo? (Pregunta descriptiva)

(Continúa en la página siguiente)

(Continuación recuadro 6.2)

4. ¿Disminuyó la proporción de niños que contrajo sarampión como resultado del programa? (Pregunta de causa y efecto)
5. ¿Hubo una reducción en la mortalidad infantil debida a complicaciones relacionadas con el sarampión como resultado de este programa? (Pregunta de causa y efecto)

Introducción de plan de estudios basado en el mercado

Intervención: tres colegios de secundaria en tres ciudades implementan un plan de estudios basado en el mercado.

Preguntas de evaluación:

1. ¿Cuán diferente es el plan de estudios del utilizado por los colegios no participantes? (Pregunta descriptiva).
2. ¿El plan de estudios se basó en el mercado, como se requería? (Pregunta normativa)
3. ¿Hasta qué punto obtienen los graduados de estos colegios trabajos con altos salarios? (Pregunta descriptiva)
4. Como resultado de la intervención, ¿están los graduados de los colegios que utilizan el plan de estudios basado en el mercado obteniendo trabajo con mayor salario del que hubiesen obtenido de otra forma? (Pregunta de causa y efecto)

Fuente: los autores.

las relaciones postuladas en los modelos de la teoría del cambio de que, debido a una intervención, se obtuvieron ganancias que no se habrían logrado de otra forma, son de causa y efecto. Las preguntas sobre si estas ganancias producen otros efectos directos u otros impactos intermedios son también de causa y efecto.

Con frecuencia, las preguntas sobre cambios en los efectos directos son descriptivas, o bien de causa y efecto deficientemente planteadas. Si la intención es que fuesen de causa y efecto, deben plantearse nuevamente para indicar que la pregunta no es sólo sobre qué cambio ocurrió, sino también sobre si el cambio puede atribuirse a la intervención (esto es, el cambio es mayor de lo que habría sido de haber faltado la intervención). Pueden también atribuirse otros efectos directos e impactos a la intervención, según lo postula la teoría del cambio.

Identificar y seleccionar las preguntas

¿Cómo decide el evaluador cuál de las muchas preguntas potenciales plantear? Cronbach (1982) sugiere usar dos fases para identificar y seleccionar las preguntas: la fase divergente y la convergente.

La fase divergente

En la fase divergente, el evaluador desarrolla una lista exhaustiva de preguntas y preocupaciones potencialmente importantes. Pocas preguntas son eliminadas y se consultan muchas fuentes. Cronbach (1982) resume la fase divergente de la planeación de una evaluación, así:

> "El primer paso es abrir la mente a las preguntas que se van a introducir al menos brevemente como prospectos para la investigación. Esta fase constituye un acto evaluativo en sí mismo y requiere recopilación de datos, análisis razonado y juicio. Muy poca de esta información y análisis es cuantitativa. Los datos provienen de conversaciones informales, observaciones casuales y revisión de registros existentes. Los métodos naturalistas y cualitativos son particularmente adecuados para este trabajo porque, al atender las percepciones de los participantes y partes interesadas, permiten al evaluador identificar esperanzas y temores que pueden no haber salido a la superficie como asuntos de políticas…
>
> El evaluador debería intentar ver el programa con los ojos de los distintos sectores de la comunidad que toma las decisiones, incluyendo los profesionales que operarían el programa de adoptarse éste, y los ciudadanos a los cuales va a servir".

Fase divergente: Fase de una evaluación en la cual el evaluador desarrolla una lista exhaustiva de preguntas y preocupaciones potencialmente importantes.

En algún punto, no se generan ya nuevas preguntas. En ese momento, el evaluador debe detenerse y examinar la lista de preguntas y empezar a organizarlas.

La fase convergente

En la fase convergente, el evaluador estrecha la lista de preguntas generadas en la fase divergente, a fin de identificar las preguntas más críticas. ¿Cómo decide el evaluador cuáles son las preguntas más críticas? Fitzpatrick, Sanders y Worthen (2004) proponen los criterios siguientes para determinar cuáles preguntas de evaluación deben investigarse:

Fase convergente: Fase de una evaluación en la cual el evaluador estrecha la lista de preguntas generadas en la fase divergente, a fin de identificar las preguntas más críticas.

- ¿Quién utilizaría la información? ¿Quién desea saber? ¿Quién estaría contrariado si no se incluyera esta pregunta?
- ¿Reduciría la incertidumbre una respuesta a la pregunta o suministraría información no disponible fácilmente?
- ¿Produciría información importante la respuesta a la pregunta? ¿Tendría impacto en el curso de los eventos?
- ¿Esta es pregunta sólo de interés pasajero para alguien, o se centra en dimensiones críticas de interés continuado?
- ¿Se limitaría seriamente el alcance o la amplitud de la evaluación si no se incluyese esta pregunta?
- ¿Es factible responder esta pregunta dados los recursos financieros y humanos, el tiempo, los métodos y la tecnología?

Esta lista de criterios puede colocarse en una matriz para ayudar al evaluador y al cliente a estrechar la lista original de preguntas, y convertirla en un conjunto manejable (cuadro 6.2).

El evaluador debe prestar atención particular a las preguntas que planteen el cliente y las partes interesadas clave. Si existen desacuerdos sobre las preguntas, es usualmente importante resolverlos en esta etapa temprana. El proceso ayuda al evaluador y al cliente, y también a las demás partes interesadas, a establecer un sentido de propiedad compartida o asociación que puede ser valioso durante las etapas posteriores de la evaluación.

Desarrollo de preguntas buenas

Para desarrollar preguntas de evaluación buenas, el evaluador empieza identificando los problemas importantes que atiende el proyecto, el programa o la política. Como ya se anotó, los problemas principales se identifican generalmente mediante una revisión de la literatura relacionada, incluyendo evaluaciones de programas similares, la teoría del cambio y los documentos del programa, como también mediante discusiones con las partes interesadas del programa y la financiación de las evaluaciones por parte del cliente.

En los ejemplos de problemas principales que debe atender una evaluación de un programa dirigido a reducir la mortalidad infantil se incluyen los siguientes:

- Causas múltiples de mortalidad infantil.
- Programas competidores en curso.
- Eficacia de los métodos utilizados para llegar a las madres de ingresos bajos.
- Grado y naturaleza de uso de suplementos alimenticios para propósitos no previstos.

Cuadro 6.2 *Matriz para seleccionar y clasificar las preguntas de la evaluación*

	Pregunta de evaluación									
La pregunta de evaluación:	1	2	3	4	5	6	7	8	9	10
¿Sería de interés para audiencias clave?										
¿Reduciría la incertidumbre actual?										
¿Produciría información importante?										
¿Sería de interés continuo (no transitorio)?										
¿Sería crítica para el alcance y la integridad del estudio?										
¿Tendría impacto en el curso de los hechos?										
¿Admitiría respuesta dados los recursos financieros y humanos, el tiempo, los métodos y la tecnología disponibles?										

Fuente: adaptado de Fitzpatrick, Sanders y Worthen, 2004.

Una vez identificados estos problemas, el evaluador puede hacer las preguntas que ayuden a determinar si los problemas han sido afectados por la política o la intervención. En las muestras de preguntas para conocimiento de los problemas se incluyen las siguientes:

- ¿Qué métodos de actividades de extensión se han utilizado en el programa?
- ¿Qué métodos de actividades de extensión han sido más eficaces?
- ¿Cuál fue la incidencia de las enfermedades infantiles graves para la vida durante el tiempo en que el programa fue operativo?
- ¿En cuánto se ha reducido la tasa de mortalidad?
- ¿Qué otros esfuerzos relacionados para mejorar la salud materna se han efectuado?

En las preguntas en las que se incluye más de un asunto –por ejemplo, "¿cuántas mujeres han recibido pruebas de tamizado de salud y suplementos nutritivos?"– deben evitarse, y en su lugar separarse la pregunta en dos ("¿Cuántas mujeres han recibido pruebas de tamizado de salud? ¿Cuántas mujeres han recibido suplementos nutritivos?").

Las preguntas sobre un asunto pueden tratarse utilizando los tres tipos de preguntas, ajustando la redacción. En una evaluación de un programa dirigido a reducir las heridas y muertes por minas terrestres, por ejemplo, se podrían hacer las preguntas siguientes:

- ¿Dónde ocurre la mayoría del minado terrestre? (Descriptiva)
- ¿Logró el proyecto la meta de eliminar 1.000 minas terrestres en la zona en el tiempo dado? (Normativa)
- ¿Ha disminuido el número de heridos o de muertos por minas terrestres como resultado de la intervención? (De causa y efecto)

Las sugerencias siguientes pueden ayudar a los evaluadores a redactar preguntas mejores:

- Establecer un vínculo claro entre cada una de las preguntas de la evaluación y el propósito del estudio.
- Asegurarse de que las preguntas de evaluación vayan dirigidas a los problemas de mayor preocupación.
- Asegurarse de que todas las preguntas admitan respuesta.
- Establecer un número realista de preguntas.
- Centrarse en las preguntas importantes, las que deben responderse, en oposición a las que sería bueno investigar.
- Asegurarse de que las preguntas admitan respuesta dados el marco de tiempo y los recursos disponibles de la evaluación.
- Considerar el momento de la evaluación en relación con el ciclo del programa. Por ejemplo, las preguntas sobre impacto se responden mejor después que la intervención haya estado operando durante varios años.

001El camino.indd 218 08/11/2010 03:27:49 p.m.

Las preguntas de la evaluación pueden relacionarse con un proyecto, un programa, un asunto de política predominante, o una intervención específica asociada con una política. Por ejemplo, si la preocupación general (el asunto de política) es reducir la pobreza, pueden lanzarse varias intervenciones de programas. Cada política se traduce a la acción mediante una intervención diseñada para lograr objetivos específicos. En última instancia, si la política y las intervenciones se realizan con eficacia y la teoría del cambio es correcta, deben lograrse los efectos directos generales. Si no es así, tanto las intervenciones como la política deben reevaluarse, y una o ambas deberán modificarse.

Diseño de la evaluación

En gran medida, así como un arquitecto diseña un edificio, un evaluador diseña una evaluación. Un diseño de evaluación es el plan de lo que se incluirá en la evaluación. No es el plan completo de trabajo para el estudio.

Un diseño de evaluación consiste en:

- Planteamiento del asunto o la pregunta principal de la evaluación.
- Selección del enfoque general de la evaluación.
- Preguntas y subpreguntas de evaluación específicas.
- Operativización (medidas o indicadores), fuentes de datos y estrategias metodológicas para la clase de recopilación de datos que se usará.
- Planeación del análisis.
- La difusión de la estrategia.

Patton (1997) distingue dos clases de cuestiones de diseño: las conceptuales y las técnicas. Las cuestiones conceptuales se centran en qué piensan las personas participantes sobre la evaluación. En ellas se incluyen temas como determinar el propósito de la evaluación y sus partes interesadas primarias y temas políticos que deben tenerse en cuenta.

Las cuestiones técnicas se refieren al plan de recopilación y análisis de datos. Estas cuestiones técnicas son el corazón de la matriz de diseño que debe desarrollarse para cualquier evaluación. Para cada pregunta, o más usualmente, subpregunta, la matriz de diseño requiere:

- Determinar la clase de pregunta o subpregunta que se hace (descriptiva, normativa, de causa y efecto).
- Especificar la medida (indicador o variable) por la cual se responderá la pregunta o subpregunta (por ejemplo, crecimiento porcentual en vivienda local o número de niños vacunados).
- Identificar el diseño metodológico que suministrará la información apropiada para responder la pregunta descriptiva, normativa o de causa y efecto.
- Identificar los recursos de datos para cada pregunta o subpregunta.
- Determinar si es necesario un marco de muestra y, de ser así, qué clase se utilizará.

- Identificar la clase de instrumentos de recopilación de datos que se utilizarán para cada pregunta o subpregunta.
- Identificar cómo se analizarán y presentarán los datos.

A veces, la medida por la cual se conteste la pregunta es un indicador acordado con un objetivo claro y la fecha para la cual deberá lograrse. Esta situación es la ideal; es el caso más frecuente cuando se ha desarrollado un marco de seguimiento y evaluación para la intervención o ésta forma parte de un sistema de seguimiento más amplio para un sector o un ministerio del gobierno, por ejemplo. Sea que se fije o no el objetivo, debe indicarse la presencia o la falta de una medida básica.

La matriz de evaluación completa representa el diseño de la evaluación, aunque no es el plan de trabajo completo: no indica todas las tareas ni identifica quién llevará a cabo cada una ni cuándo. El plan de trabajo completo se cubre en el capítulo 12.

Etapas en el proceso de diseño de la evaluación

Idealmente, el proceso de evaluación empieza *ex ante*, con el diseño inicial del programa, y luego prosigue en varias etapas distintas e importantes.

Etapa 1. Planear o establecer el alcance de la evaluación

La fase inicial de planeación o de fijación del alcance clarifica la naturaleza y alcance de la evaluación. Durante esta fase se establecen el propósito principal de la evaluación, las partes interesadas a quienes se les consultará, la persona que realizará la evaluación y el marco de tiempo para los resultados. Este período es exploratorio. Se identifican los temas clave desde la perspectiva del cliente principal y otras partes interesadas, la revisión de la literatura y las intervenciones relacionadas que puedan influir en el programa. Se desarrollan la teoría del cambio y sus supuestos fundamentales.

Etapa 2. Diseñar la evaluación

Al finalizar la fase inicial de planeación o de fijación de alcance, debe haber suficiente conocimiento del contexto, de modo que pueda decidirse el enfoque general para la evaluación. El centro de la planeación de la evaluación es la fase de su diseño, que culmina en la matriz de diseño de la evaluación. Un diseño general deficiente limitará la capacidad de extraer conclusiones sobre el desempeño de la intervención.

En general, una buena práctica es presentar y discutir el diseño general con el patrocinador (cliente) de la evaluación y otras partes interesadas clave, antes de finalizar el diseño de la evaluación. Hacerlo así asegura que no habrá sorpresas y genera aceptación y apoyo para la evaluación. Un grupo asesor y de revisores pares constituyen también juntas de sondeo para asegurar la solidez del diseño de la evaluación. En casos de alto perfil, las propuestas de diseño pueden colocarse en un sitio web para comentarios.

001El camino.indd 220 08/11/2010 03:27:49 p.m.

La matriz de diseño puede utilizarse como base para desarrollar los términos de referencia (TDR), que sirven de base para una solicitud de propuesta o como guía para el equipo de evaluación si la evaluación se efectúa internamente. Cuando el trabajo de fijación de alcance y de antecedentes para la evaluación los va a realizar un consultor externo, es aconsejable poner la matriz como un producto de entrega para este trabajo. Posteriormente, otros TDR pueden desarrollarse para implementar el diseño de evaluación.

Etapa 3. Realizar la evaluación

La "fase de realización" de la evaluación incluye la recopilación y el análisis de los datos. Usualmente, si se van a recopilar distintas clases de datos (o datos similares de distintas fuentes), deben desarrollarse y probarse diferentes instrumentos. El análisis se hace a menudo en forma concurrente con la recopilación de datos.

Después de poco más o menos las dos terceras partes del trabajo de recopilación de datos, el equipo de evaluación debería sostener una conferencia de validación para examinar los hallazgos hasta la fecha e identificar los temas y mensajes principales. Una conferencia de validación es una forma útil de llegar a un acuerdo temprano sobre los tres a cinco mensajes principales. El propósito de la conferencia de validación es asegurar el acuerdo temprano sobre los temas principales y verificar que la cuestión principal subyacente en la evaluación hayan sido atendidos. (Aunque el esbozo del informe puede haberse organizado alrededor de las preguntas de la evaluación, puede ser más eficaz organizar el informe definitivo y comunicarse con quienes toman las decisiones por mensaje o tema, pues no es probable que todas las preguntas de la evaluación tengan el mismo interés.)

Conferencia de validación: Reunión en la cual el equipo de evaluación discute y se pone de acuerdo con un tema principal, y verifica que el asunto principal subyacente en la evaluación haya sido atendido.

Etapa 4. Informar los hallazgos de la evaluación

En la fase de preparación y presentación del informe, los hallazgos o los enunciados iniciales de hecho se pueden compartir y discutir con los "propietarios" del programa, de modo que pueda corregirse cualquier error factual y cualquier información novedosa pueda tenerse en consideración antes del borrador del informe y el desarrollo de las recomendaciones. Una vez finalizado el análisis, se preparan los resultados, se revisan los borradores, se incorporan los comentarios apropiados y se presenta el informe final al cliente y las partes interesadas clave.

Un informe usualmente proporciona los antecedentes, y el contexto para la evaluación, indica el propósito de ésta e informa sobre los hallazgos (incluyendo los efectos directos previstos y los imprevistos).

Generalmente también incluye información sobre las enseñanzas aprendidas y las recomendaciones. Comprender lo que no funciona bien y por qué es tan importante

001El camino.indd 221 08/11/2010 03:27:49 p.m.

como comprender lo que sí funciona y por qué, y ambas cosas deben estar claras. El informe debe escribirse teniendo presente la audiencia a la cual va dirigido, y debería estar libre de jerga y ser de fácil lectura. (La escritura del informe se trata en el capítulo 13.)

Etapa 5. Difusión y seguimiento de los hallazgos de la evaluación

Planear la evaluación significa planear para la comunicación durante el proceso, no sólo con el cliente y las partes interesadas clave, sino también dentro del equipo de evaluación. Una evaluación no está completa sino hasta que la difusión se realice; por consiguiente, el desarrollo de un plan de difusión forma parte del proceso de planeación. No siempre es necesario presentar los hallazgos en forma impresa. Los resúmenes son especialmente útiles para comunicar los hallazgos mientras la evaluación está en curso, especialmente cuando los hallazgos son inesperados o críticos.

Los resultados de muchas evaluaciones originan acciones para:

- Modificar una intervención.
- Eliminar barreras identificadas en la evaluación.
- Informar sobre políticas o intervenciones futuras (modificar la teoría del cambio).
- Mostrar a otros el camino en relación con las enseñanzas aprendidas.
- Rediseñar el pensamiento sobre la naturaleza del problema.

Muchas organizaciones tienen sistemas de seguimiento para rastrear las recomendaciones formales y resumir las enseñanzas. La capacidad de indagar en tales bases de datos por tema, sector, ubicación y fecha aumenta su utilidad. Ya sea que se encuentren en funcionamiento tales sistemas o no, los evaluadores deben considerar enviar una copia electrónica no editable del informe a las bases de conocimiento sobre evaluación pertinentes. También pueden considerar presentar sus hallazgos en conferencias sobre evaluación, o enviar un artículo sobre la evaluación para su publicación en una revista profesional.

Relaciones entre las etapas

En el recuadro 6.3 se resumen las distintas etapas, y en el gráfico 6.2 se presenta la relación entre los distintos componentes.

En el gráfico 6.2 se ilustra la centralidad de promover el uso de la evaluación al proceso de evaluación completo. En lugar de ser algo en lo que se piensa al finalizar la evaluación, promover el uso de la evaluación es el centro de ésta, que orienta a otros procesos o etapas.

La matriz de diseño de la evaluación

Una matriz de diseño de evaluación es una herramienta organizativa muy recomendada para planear una evaluación. En ella se organizan las preguntas y se planea la recopilación de información para contestarlas. La matriz vincula las preguntas descriptivas,

Recuadro 6.3 *Las cinco etapas del proceso de la evaluación*

Etapa 1. Planear o establecer el alcance de la evaluación

Obtener una comprensión completa del proyecto, programa o política.

- Reunirse con el usuario principal para la evaluación.
- Identificar y reunirse con otras partes interesadas clave.
- Explorar el contexto del programa y acumular materiales de antecedentes.
- Buscar evaluaciones pertinentes relacionadas.
- Revisar las evaluaciones anteriores para identificar temas, diseños y estrategias de recopilación de datos utilizadas.
- Reunirse con el personal del programa (si es externo).
- Revisar y refinar o desarrollar una teoría del cambio para el programa.

Etapa 2. Diseñar la evaluación

Determinar las preguntas y problemas

- Reunirse con el usuario, e identificar el propósito principal de la evaluación, los temas de interés y las necesidades de tiempo críticas.
- Identificar y reunirse con otras partes interesadas clave, a fin de identificar problemas y preocupaciones para su posible inclusión en la evaluación.
- Determinar los recursos disponibles para la evaluación, por ejemplo, presupuesto para consultores y viajes, integrantes de los equipos y mezcla de destrezas.
- Evaluar las necesidades de las partes interesadas, incluyendo el tiempo.

Preparar los términos de referencia y la matriz de evaluación.

- Identificar el tipo de evaluación.
- Identificar las preguntas y subpreguntas específicas de la evaluación.
- Seleccionar medidas para cada pregunta o subpregunta.
- Identificar las fuentes de datos para tratar cada pregunta o subpregunta.
- Identificar un diseño apropiado para cada pregunta o subpregunta.
- Desarrollar una estrategia de recopilación de datos, incluyendo los instrumentos y métodos de muestreo que se utilizarán para cada pregunta o subpregunta.
- Desarrollar una estrategia para analizar los datos.
- Determinar los requerimientos de recursos y tiempo.

Etapa 3. Realizar la evaluación

- Informar brevemente al usuario y las partes interesadas clave sobre el diseño de la evaluación.
- Preparar un plan de trabajo, incluyendo la revisión y la prueba de la metodología, los instrumentos para las pruebas, la capacitación a los recopiladores de datos y el desarrollo del protocolo.
- Recopilar los datos.

(Continúa en la página siguiente)

(Continuación recuadro 6.2)

- Preparar los datos para el análisis, desarrollando capas (*shells*) de cuadros (si no se realizan como parte del diseño de evaluación) y limpieza de los datos.
- Analizar los datos.
- Desarrollar gráficos.
- Formular los hallazgos.

Etapa 4. Informar los hallazgos de la evaluación

- Sostener una Conferencia de validación.
- Identificar los hallazgos y temas principales: qué funciona, qué no funciona y qué necesita mejoramiento.
- Escribir el informe.
- Informar brevemente al cliente sobre los hallazgos y los enunciados de hecho.
- Informar brevemente a los funcionarios del programa y partes interesadas clave sobre los hallazgos y enunciados de hecho, y hacer las correcciones necesarias.
- Permitir a los funcionarios del programa revisar y comentar sobre el borrador del informe.
- Desarrollar recomendaciones claras y específicas e indicar quién debería hacer qué y cuándo.
- Verificar que las recomendaciones estén ligadas a la evidencia.

Etapa 5. Difusión y seguimiento de los hallazgos de la evaluación

- Determinar quién recibirá qué clase de producto de difusión del estudio (por ejemplo, un resumen, un resumen de dos a cuatro páginas, el informe completo, un taller en profundidad) e implementar el plan.
- Identificar las enseñanzas y los mecanismos para compartirlos y retenerlos.
- Hacer seguimiento a las recomendaciones formales para determinar la implementación.
- Colocar un archivo electrónico en forma no editable en los repositorios de conocimiento sobre evaluaciones.
- Considerar la difusión adicional a través de organizaciones y revistas profesionales.

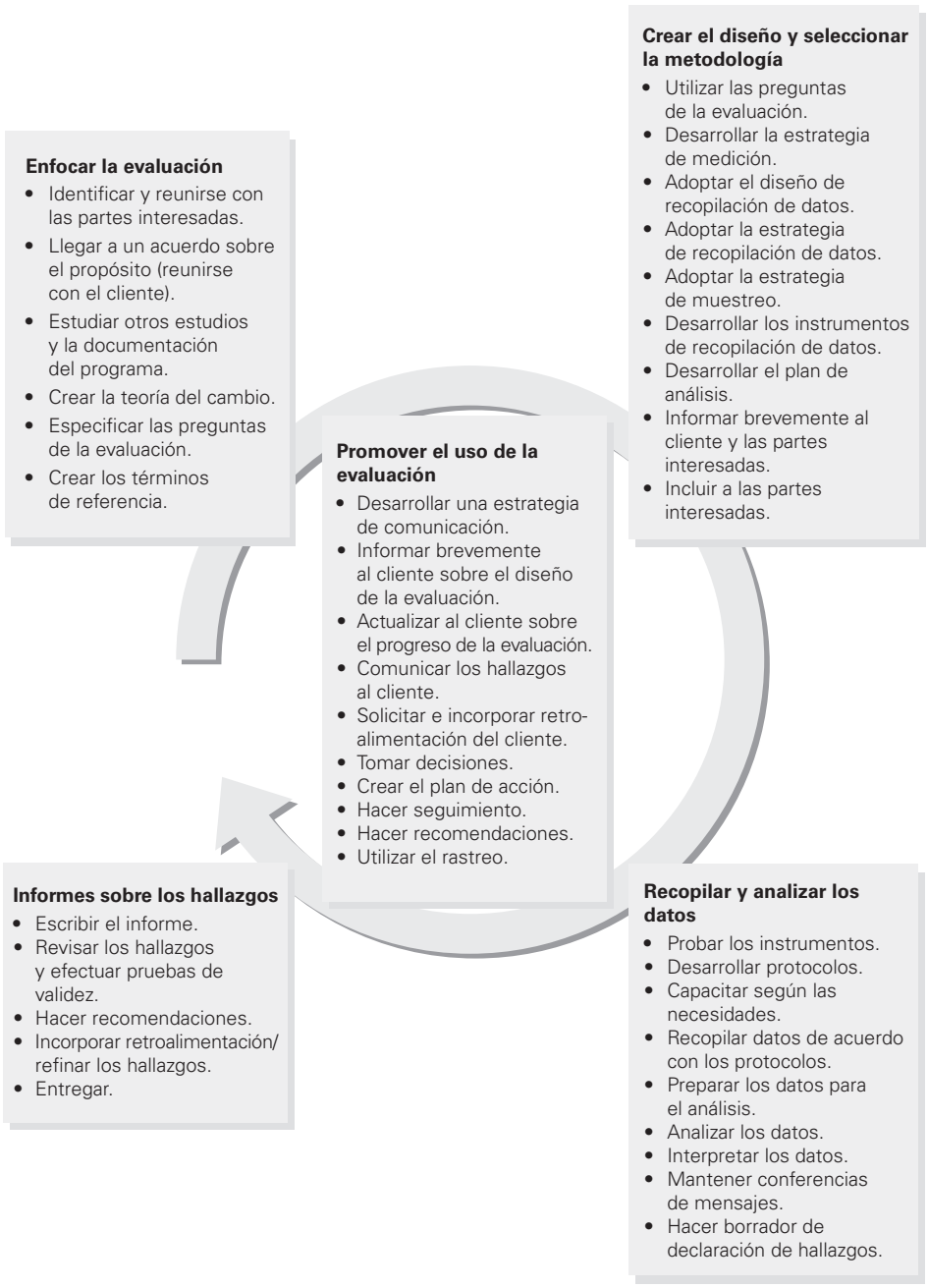
Fuente: los autores.

Matriz de diseño de evaluación: Matriz en la que se organizan las preguntas de la evaluación y se planea la recopilación de información para contestarlas.

normativas y de causa y efecto con el diseño y las metodologías. Aparte de su utilidad inmediata como herramienta de planeación, la matriz contribuye a promover el uso de la evaluación y mejorar la cooperación entre los evaluadores y el personal del programa.

Los evaluadores necesitan una herramienta para identificar las piezas necesarias de la evaluación, y asegurar que se conectan claramente en cada paso. Variará la herramienta que utilicen para pensar sobre el programa, su contexto, objetivos mensurables, recopilación y análisis de datos y estrategias que se utilizarán. Algunos evaluadores pueden decidir crear sus propias herramientas.

Gráfico 6.2 *Enfoque a la evaluación para el desarrollo*



Fuente: los autores.

El fin de la matriz de diseño es organizar el propósito y las preguntas de la evaluación, y emparejar lo que se va a evaluar con las técnicas apropiadas de recopilación de datos. En una matriz de diseño se incluyen usualmente los siguientes elementos vinculados:

- Tema principal de la evaluación.
- Enfoque general.
- Preguntas y subpreguntas.
- Tipos de preguntas y subpreguntas.
- Medidas o indicadores.
- Objetivos o estándares (si normativos).
- Presencia o ausencia de datos básicos.
- Estrategia de diseño.
- Fuentes de datos.
- Muestra o censo.
- Instrumento de recopilación de datos.
- Análisis de datos y gráficos.
- Comentarios.

Los protocolos de recopilación de datos y la evaluación sobre asignaciones de trabajos y su programación, los términos de referencia y los planes de comunicación se pueden agregar o mantener como herramientas separadas pero vinculadas.

El método de recopilación de datos puede atender una pregunta o más de una, o utilizarse varios métodos de recopilación de datos para atender una sola pregunta. La matriz de diseño incorpora fuentes de información conocidas y planeadas. Cuando el proceso pasa de la planeación a la implementación, las fuentes se pueden expandir y clarificar.

Resumen

Los evaluadores necesitan trabajar con el cliente principal y las partes interesadas clave, a fin de identificar las posibles preguntas. Después de terminar la investigación de antecedentes y reunirse con el cliente y las partes interesadas clave, desarrollar la teoría programática del cambio e identificar los supuestos principales subyacentes al programa, el evaluador puede empezar a generar y luego seleccionar las preguntas de evaluación de la larga lista de aquellas que se generaron. Las preguntas de la evaluación deben verificarse frente al tema principal de evaluación para confirmar que se está atendiendo.

Los evaluadores utilizan preguntas descriptivas, normativas y de causa y efecto. La redacción de cada pregunta es importante, porque contribuye a determinar los medios de encontrar las respuestas a la pregunta.

La forma recomendada de organizar la evaluación es utilizar una matriz de diseño. Ésta ayuda a organizar las preguntas, los diseños y las estrategias de recopilación y análisis de datos, entre otras cosas. Los capítulos siguientes ofrecen una guía paso a paso para completar la matriz de diseño.

001El camino.indd 227

Gráfico 6.3 *Plantilla de la matriz de diseño*

Matriz de diseño para: _____

Principal tema de evaluación: _____ Enfoque general de evaluación: _____

Pregunta	*Subpregunta*	*Tipo de subpregunta*	*Medida o indicador*	*Objetivo o estándar (normativo)*	*¿Datos básicos?*	*Fuente de datos*	*Diseño*	*Muestra o censo*	*Instrumento de recopilación de datos*	*Análisis de datos*	*Comentarios*

08/11/2010 03:27:50 p.m.

Actividades del capítulo 6

Ejercicio de aplicación 6.1: tipos de preguntas

Identificar si cada una de las preguntas siguientes sobre una iniciativa de salud preventiva para las mujeres rurales es una pregunta descriptiva, normativa o de causa y efecto. Si deben volver a escribirse algunas preguntas para que su tipo sea más claro, hacerlo (esto sucede a menudo en la vida real).

1. ¿Proporcionó la iniciativa la asesoría, el apoyo y otros servicios requeridos a 30 mujeres rurales en su primer mes de operaciones, según estaba planeado?
2. ¿Se prestaron los servicios en el lugar y el tiempo que maximizaba el número de mujeres que podía participar?
3. ¿Cuáles fueron los mejores métodos para llegar a las mujeres en zonas remotas y hacer accesible el programa para ellas?
4. ¿Se detectaron problemas de salud en las mujeres rurales más temprano entre quienes participaron en la iniciativa de salud para las mujeres?
5. Desde el inicio del programa, ¿cuántas mujeres han recibido qué tipo de servicios?
6. ¿Qué tan eficaz es la iniciativa de salud para mujeres, comparada con otras intervenciones para mejorar la salud de las mujeres rurales?
7. ¿Cuál es el impacto de la iniciativa de salud en las mujeres, sus familias y la comunidad rural general en la que viven?
8. ¿Cuán satisfechas están las participantes con la asesoría, la información, el apoyo y otros servicios que reciben?
9. ¿Está cumpliendo la iniciativa de salud para mujeres rurales los estándares de eficiencia requeridos por el gobierno?
10. ¿Cuáles dicen las participantes que son los impactos del programa en ellas?
11. ¿Hasta qué punto las mujeres que recibieron los servicios cumplieron los requerimientos de elegibilidad?
12. ¿Cumplió el programa su objetivo de aumentar el conocimiento de las mujeres sobre técnicas preventivas?

Ejercicio de aplicación 6.2: modificar los tipos de preguntas

Escriba una pregunta descriptiva, una normativa y una de causa y efecto para cada uno de los programas siguientes:

1. Un programa de capacitación vocacional para jóvenes varones.
2. Un programa de construcción de vías que vincula tres comunidades con un mercado central.

Referencias y lecturas adicionales

Cronbach, L. J. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.

Feuerstein, M. T. 1986. *Partners in Evaluation: Evaluating Development and Community Programs with Participants*. Londres: MacMillan, en asociación con Teaching Aids at Low Cost.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines*. Nueva York: Pearson Education Inc.

Human Rights Resource Center. 2000. "Questions about Evaluation". En *The Human Rights Education Handbook: Effective Practices for Learning, Action, and Change*. University of Minnesota, Minneápolis. http://www1.umn.edu/humanrts/edumat/hreduseries/hrhandbook/part6B.html.

Lawrence, J. 1989. "Engaging Recipients in Development Evaluation: The 'Stakeholder' Approach". *Evaluation Review* 13 (3): 243-56.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2007. *Criteria for Evaluating Development Assistance*. Development Assistance Committee. http://www.oecd.org/document/22/0,2340,en_2649_201185_2086550_1_1_1_1,00.html.

Patton, Michael Quinn. 1997. *Utilization-Focused Evaluation*. 3ª ed. Thousand Oaks, CA: Sage Publications.

———. 2002. *Qualitative Research and Evaluation Methods*. 3ª ed. Thousand Oaks, CA: Sage Publications.

Rist, R. C. 1994. "Influencing the Policy Process with Qualitative Research". En *Handbook of Qualitative Research*, ed. N. K. Denzin and Y. S. Lincoln, 545-57. Thousand Oaks, CA: Sage Publications.

Shadish, William 1998. "Some Evaluation Questions". *Practical Assessment, Research & Evaluation* 63. http://PAREonline.net/getvn.asp?v=6&n=3. También disponible en http://www.ericdigests.org/1999-2/some.htm.

Sitios web

Banco Mundial. 1996. *Participation Sourcebook*. http://www.worldbank.org/wbi/sourcebook/sbhome.htm.

Kellogg Foundation. 2004. *Evaluation Logic Model Development Guide*. http://www.wkkf.org/pubs/tools/evaluation/pub3669.pdf.

001El camino.indd 230 08/11/2010 03:27:50 p.m.

CAPÍTULO 7

Selección de diseño para preguntas de evaluación de causa y efecto, normativas y descriptivas

Después de escoger las preguntas de evaluación, el evaluador selecciona el enfoque de diseño de evaluación que sea más apropiado según cada pregunta. En este capítulo se presentan algunos lineamientos sobre el diseño y se identifican las fortalezas y las debilidades de varias opciones de diseño. Es importante tener presente, sin embargo, que cada situación es única. No existe una elección perfecta de diseño y, por tanto, tampoco una forma "sola y única" de tratar una pregunta de evaluación.

Conectar las preguntas al diseño

Al evaluar se busca responder preguntas. En el capítulo 6 se indicó que había tres tipos principales de preguntas: descriptivas, normativas y de causa y efecto. El diseño seleccionado debería ser apropiado para el tipo de pregunta que se hace. Si, por ejemplo, se plantean preguntas de causa y efecto, debe adoptarse un diseño que pueda responder preguntas de ese tipo.

Las organizaciones para el desarrollo buscan encontrar soluciones a las preguntas sobre temas de desarrollo, así como los aldeanos podrían buscar aprender a

resolver un problema que acosa a su aldea (recuadro 7.1). Sin embargo, al intentar responder las preguntas, ni los evaluadores ni los aldeanos han dado siempre los pasos correctos.

El primer paso en falso potencial al responder una pregunta es empezar escogiendo una estrategia para recopilar los datos. Para responder la pregunta de la joven sobre los elefantes, el anciano de la aldea podría decir: "Realicemos una encuesta y averigüemos lo que dicen los aldeanos que hace que los elefantes se vayan". Casi con toda certeza, el empezar con una estrategia de recopilación de datos va a suministrar la información que se requiere.

El segundo paso en falso potencial es pensar que cada evaluación tiene un solo diseño. Usualmente, una evaluación busca atender varias preguntas, cada una de ellas requiere un diseño apropiado. Una evaluación requerirá usualmente atender preguntas descriptivas, normativas y algunas veces preguntas de causa y efecto. El evaluador debe evitar aplicar la técnica del "método en busca de una aplicación" (¿qué preguntas pueden responderse realizando una encuesta o grupos de discusión?), o pensar que si se atiende una pregunta de causa y efecto, no hay por qué atender preguntas descriptivas y normativas.

En el ejemplo del elefante, es posible que un estudio de caso en profundidad de los movimientos de los elefantes muestre que ni el golpeteo de las ollas ni el polvo hagan que los elefantes salgan de la aldea. Simplemente puede ser que la aldea esté en la ruta migratoria de los elefantes.

Categorías generales de diseño

Los evaluadores tienen tres categorías generales de diseño para escoger: diseños experimentales, diseños cuasiexperimentales y diseños no experimentales. Cada uno de ellos se describe a continuación.

Recuadro 7.1 *¿Qué hace irse a los elefantes?*

Dos veces al año, desde direcciones opuestas, los elefantes pasaban rabiosos a través de una aldea africana. Todos los aldeanos se ocuparon de hacer salir a los elefantes. Algunos golpearon ruidosamente ollas y cacerolas, otros silbaron, gritaron o dieron alaridos. Otros levantaron polvo pateando el suelo y se movieron en todas direcciones como un esfuerzo para establecer su propiedad de la tierra.

Después que los elefantes salieron de la aldea, una joven preguntó: "Por qué se fueron los elefantes?" Los aldeanos respondieron: "Porque nosotros los hicimos huir". Entonces la muchacha preguntó: "Pero, ¿qué los hizo huir, el ruido del golpeteo de las ollas, los silbidos, otros ruidos altos, o el polvo en el aire?"

Diseño experimental

Diseño experimental: Tipo de diseño de evaluación *ex ante* que requiere la asignación aleatoria de una población a por lo menos dos grupos, tales que cada miembro de la población de interés tenga igual probabilidad de ser asignado a una intervención (grupo de tratamiento), o a un grupo externo a la intervención (grupo de control).

Muchos evaluadores consideran un diseño experimental –llamado también *experimento al azar* o *experimento verdadero*– como el diseño más fuerte y más robusto para una evaluación, aunque muchos otros están en desacuerdo con esto. Como se argumentó en el capítulo 6, los evaluadores deben esforzarse por ser tan rigurosos como sea posible, pero el diseño correcto es el más apropiado para responder las preguntas de la evaluación. En un verdadero diseño experimental, los evaluadores deben mostrar que de no haber ocurrido la intervención, no se habría logrado el resultado deseado. Utilizando a los elefantes del ejemplo de la aldea, el evaluador debe demostrar que si los aldeanos no hubiesen golpeteado las ollas, silbado y hecho otros ruidos, los elefantes no habrían salido de la aldea.

Para mostrar que la intervención es la causa de un efecto directo, en un diseño experimental se comparan los resultados de dos grupos: uno que recibe una intervención y otro que no la recibe. El criterio principal para distinguir diseños experimentales de otros diseños es la asignación aleatoria a los grupos. Mediante la asignación aleatoria, los asignados a los grupos deben ser similares en cuanto a antecedentes, género, contexto, tiempo, etc.

Aquí se interrumpe nuestra analogía de los elefantes, porque los aldeanos no tienen forma de separar a los elefantes en dos o tres grupos asignados al azar, que pudiesen estar sujetos a distintas intervenciones o a ninguna. Pero para continuar con el ejemplo, podría decirse que el gobierno nacional tiene conocimiento de que 30 aldeas de la provincia rural han informado sobre problemas con los elefantes que causan destrucción en los hogares y cultivos de las aldeas. El gobierno quiere resolver este problema de manera que los elefantes y los aldeanos vivan en armonía. Se contrata a un experto en comportamiento animal, pero éste sólo puede trabajar con 20 aldeas a la vez. Se decide que las 20 aldeas que reciban la ayuda del experto serán seleccionadas al azar, y que las 10 aldeas no seleccionadas servirán como grupo de control. Se efectúan encuestas a los aldeanos adultos de las 30 aldeas con el fin de obtener datos básicos para los grupos de tratamiento y de control sobre las actitudes hacia los elefantes y estimados de pérdidas de cultivos y otros daños. Para la selección de las aldeas que recibirán la intervención, se escriben los nombres de las 30 aldeas en papeles, y se colocan éstos en una vasija. Se reúne a los ancianos de la aldea y se les solicita por turno a los 10 más importantes seleccionar (a ciegas) un papel de la vasija. Cuando se lee cada nombre en el papel, se escribe el nombre de la aldea en un tablero, y cuando se haya identificado a 10 aldeas, ha finalizado el grupo de control. El objetivo de la intervención conductista es reducir el problema, mejorar las actitudes y reducir los daños y los costos asociados a éstos. Después de un período preestablecido, deberían existir diferencias entre las aldeas de tratamiento y de control sobre estas medidas, si el experto es eficaz y todo lo demás permanece constante. En este ejemplo, el conductista es la intervención.

Diseño cuasiexperimental

Un diseño cuasiexperimental es similar a un diseño experimental, pero en el primero no se asignan al azar los individuos a los grupos. Estos diseños son a menudo más factibles que los experimentales, porque con mucha frecuencia los diseños experimentales verdaderos son difíciles de ejecutar, y los cuasiexperimentales no requieren el tratamiento al azar sino la creación de grupos de comparación.

Diseño cuasiexperimental: Tipo de diseño de evaluación en el que los grupos de tratamiento y control o comparación se forman *ex ante* o *ex post,* pero sin asignación aleatoria a los grupos, o en el que se toman medidas repetidas a través del tiempo.

Un diseño cuasiexperimental puede utilizarse para comparar grupos similares, pero no equivalentes. Los grupos pueden existir en aldeas diferentes pero similares, o utilizarse el mismo grupo en tiempos distintos (antes y después). Por ejemplo, en el caso de los elefantes de las aldeas, con un diseño cuasiexperimental podrían elegirse dos aldeas en la misma región, con el mismo clima, el mismo número de elefantes en la zona, la población similar de aldeanos y el número similar de hogares, etc. Una aldea recibiría el tratamiento (golpear las ollas) y la otra no.

Alternativamente, podrían seleccionarse los dos grupos, modificando el comportamiento de los aldeanos al llegar los elefantes. La primera vez que éstos se presenten en la aldea, los aldeanos golpean las ollas. La segunda vez no las golpean. Así se tienen dos instancias para comparar la reacción de los elefantes.

En ambos ejemplos, los dos grupos son similares, pero no equivalentes. En el primer ejemplo, las diferencias entre los dos grupos y sus entornos pueden afectar los resultados. En el segundo ejemplo, la diferencia entre las dos situaciones es el comportamiento de los aldeanos. Con un diseño cuasiexperimental, los evaluadores no pueden definitivamente vincular la intervención a la solución, o mostrar un vínculo de causa y efecto con la certeza de que produce la verdadera aleatorización, pero pueden aprender mucho y postular sobre la probable causa y el efecto.

Diseño no experimental

En los diseños no experimentales, o descriptivos, no se compara un grupo con otro. En cambio, se proporciona una descripción extensa de la relación entre una intervención y sus efectos. Con un estudio no experimental, un evaluador escoge, por ejemplo, cuándo hacer la muestra, de entre quiénes extraerla y dónde efectuarla. No se hace ningún esfuerzo por crear dos o más muestras equivalentes o similares.

Diseño no experimental (descriptivo): Tipo de diseño de evaluación en el que no se hace el intento de crear grupos de tratamiento y de control, y el énfasis se hace en la descripción.

Por ejemplo, en una evaluación no experimental puede utilizarse un análisis de los datos o la información existente, una encuesta o un grupo de discusión para acumular datos apropiados pertinentes a las preguntas de la evaluación. Los diseños no experimentales tienden a examinar características, frecuencia y asociaciones (Project STAR, 2006).

Cuadro 7.1 *Dos características clave de los diseños experimentales, cuasiexperimentales y no experimentales*

Tipo de diseño de evaluación	*Grupo de control asignado al azar*	*Grupo de comparación no aleatorio*	*Medidas repetidas*
Experimental	Sí	No	Sí
Cuasiexperimental	No	Probable	Probable
No experimental	No	No	No

Fuente: los autores.

En el cuadro 7.1 se resumen tres de las características clave de cada una de las tres categorías de diseño.

Notación de diseño

Los diseños de evaluación se representan algunas veces utilizando las letras X y O. En estas representaciones, una X representa una intervención o un tratamiento, y una O representa una observación. A cada tratamiento y observación se le asigna un subíndice para identificarla. Por ejemplo, la notación para un diseño de evaluación que tiene un tratamiento seguido de una observación es

$$X\ O_1$$

La notación para un diseño de evaluación que tiene una observación seguida del tratamiento, seguido de dos observaciones, sería

$$O_1\ X\ O_2\ O_3$$

A cada grupo en el diseño se le da una línea separada. La siguiente es la notación para un diseño de evaluación que tiene dos grupos, uno que recibió el tratamiento y uno que no. Ambos grupos se observan una vez antes del tratamiento al grupo de prueba y dos veces después:

$$O_1\ X\ O_2\ O_3$$

$$O_1\ O_2\ O_3$$

Diseños con preguntas de causa y efecto

Las preguntas de causa y efecto plantean el mayor problema, ya que responderlas requiere un diseño particularmente bien analizado. En las evaluaciones que incluyen preguntas de causa y efecto –es decir, sobre los impactos y causas de estos impactos observados– con el diseño de la evaluación se intenta descartar explicaciones factibles para resultados

observados distintos de la intervención a fin de concluir que la intervención fue la que causó el impacto. En resumen, el problema es decidir si atribuir o no los cambios observados (impactos) a la intervención o a alguna otra causa.

Cuando se atienden preguntas de causa y efecto, los diseños de evaluación pueden enmarcarse también para atender la pregunta "¿Cuál habría sido la situación si la intervención no hubiera tenido lugar?" Puede no ser posible medir con precisión el contrafactual de no intervención, pero sí estimar qué habría sucedido sin intervención.

Diseños experimentales

El modelo experimental tiene sus raíces en la investigación médica, en la que se utiliza a menudo para hacer pruebas de medicamentos y protocolos de tratamiento. Al aplicar este diseño a una pregunta de evaluación relacionada con la salud, una organización para el desarrollo que busca reducir la incidencia de la malaria en una región podría plantear la pregunta "¿Cuál es la mejor forma de reducir la incidencia de la malaria en la región?" Una subpregunta podría ser "¿Reducen los toldos para los mosquitos la incidencia de la malaria en esta región?"

El diseño experimental toma una pregunta y la convierte en una proposición. En el caso del ejemplo de la malaria, la proposición sería así: si las personas en la región reciben toldos para los mosquitos, entonces habrá menos casos de malaria en la región.

Como se presentó en el capítulo 4, debe desarrollarse una teoría del cambio para la proposición. El tratamiento al azar es el factor principal en un diseño experimental. Se puede seleccionar un grupo de pacientes entre voluntarios para pruebas aleatorias que tengan la misma etapa y mismo tipo de enfermedad, el mismo género, etc. Entonces a los individuos se les asigna al azar uno de varios tratamientos de medicamentos. Un subgrupo podría estar tomando el medicamento que representa el conocimiento actual y otro recibir un nuevo medicamento prometedor. En un experimento doble ciego (*double-blind*), ni los participantes en el estudio ni el personal médico saben qué medicamento se administra a qué pacientes.

En la evaluación para el desarrollo existe una ligera pero creciente tendencia a utilizar diseños experimentales. El movimiento en esta dirección proviene de una frustración con la percepción de falta de conocimiento suficiente, a pesar de muchos años de evaluación, de lo que funciona en el contexto del desarrollo y en qué condiciones.

Un experimento clásico tiene seis pasos:

1. Formular una hipótesis.
2. Obtener datos básicos (es decir, medir la variable dependiente).
3. Asignar al azar casos a los grupos de intervención y de control.
4. Introducir el tratamiento o la variable independiente en la intervención.
5. Medir otra vez la variable dependiente (posteriormente a la prueba).
6. Calcular las diferencias entre los grupos y someter a prueba la significación estadística.

En el ejemplo de la malaria, la pregunta es si los toldos para los mosquitos reducen la incidencia de la malaria en una región. Los seis pasos serían así:

1. *Formular una hipótesis*: el uso de toldos con repelente para mosquitos por parte de las familias reduce la incidencia de la malaria.
2. *Obtener datos básicos*: durante un período de dos meses, el número de casos nuevos de malaria en la región fue de 172.
3. *Asignar al azar casos a los grupos de la intervención y de control.*
4. *Introducir el tratamiento*: dar toldos a un grupo (el de tratamiento) y no al otro (el de control).
5. *Medir otra vez la variable dependiente*: después de dos meses con la intervención, el número de casos nuevos de malaria fue de 65 en el grupo de tratamiento y de 188 en el de control.
6. *Calcular la diferencia entre los dos grupos*: el grupo de tratamiento tuvo 53 casos menos que el de control.

La notación de diseño para el experimento clásico con un grupo de tratamiento y uno de control es semejante a ésta:

$$O_1 \; X \; O_2$$

$$O_1 \; O_2$$

En algunos casos, los evaluadores utilizan también una R en el frente de la línea para cada grupo a fin de mostrar que el grupo tuvo asignación al azar. Para un diseño experimental como el descrito antes (asignar al azar algunos participantes para recibir un toldo), la notación sería como ésta:

$$R \; O_1 \; X \; O_2$$

$$R \; O_1 \; O_2$$

Los seres humanos somos complejos y las situaciones sociales son difíciles de analizar. El diseño experimental funciona bien cuando se prueban medicamentos en los cuerpos humanos, que son relativamente similares. Al observar el complejo comportamiento humano, debe estarse alerta por falsos positivos y falsos negativos. Los falsos positivos ocurren cuando el estudio indica que la intervención fue la causa del efecto cuando en realidad no lo fue. Los falsos negativos ocurren cuando un estudio no muestra vínculo alguno entre una intervención, y el éxito cuando, en realidad, la intervención realmente está vinculada al éxito.

Un falso positivo puede aparecer debido a que los datos de las familias son auto-informados. Las personas de la región pueden saber del experimento y de las familias del tratamiento, e informar menos casos de malaria para complacer a los funcionarios del gobierno.

Los falsos negativos son más comunes. Puede no encontrarse diferencia alguna entre los dos grupos, porque el período de medición de dos meses después de la intervención fue tomado en la estación seca, cuando la incidencia de malaria es baja. Un falso negativo puede ocurrir porque las familias del grupo de tratamiento no utilizan los toldos todas las noches, o porque los del grupo de control compran los toldos por su cuenta.

Si en este estudio no se hicieran preguntas adicionales sería difícil interpretar los resultados, aun en esta intervención simple de asignación al azar. Para explicar los resultados habría sido útil que los evaluadores hubiesen hecho algunas de estas preguntas:

- ¿Qué información recibieron las familias sobre el uso del toldo para los mosquitos?
- ¿Cómo se determinó la composición de la "familia" y quién realmente durmió utilizando el toldo?
- ¿Qué problemas de implementación suscitaron los grupos de la intervención y de control?
- ¿Cómo se comparó la incidencia de la malaria en la región históricamente para los dos períodos de dos meses?
- ¿Hubo otros esfuerzos de prevención de malaria en la región en el momento de la intervención?

Todas estas preguntas son descriptivas, que requieren diseños más simples para responderlas.

Grupos de control

Con un diseño experimental se intenta descartar o controlar otros factores que pueden representar explicaciones competidoras para los resultados del experimento. Al utilizar el diseño experimental, los evaluadores comparan grupos equivalentes. El *grupo de control* es el que está expuesto a las condiciones usuales; sus miembros no están expuestos a la intervención o no se les administra. El grupo expuesto a la intervención puede llamarse el *grupo de tratamiento.* Utilizar un grupo de control permite comparar los grupos expuestos a la intervención con grupos no expuestos al tratamiento.

Grupo de control: Grupo en un experimento cuyos miembros no están expuestos a una intervención o no se les administra.

Los grupos de control implican a menudo no administrar una intervención a algunos de los que tienen necesidad. Algunas veces esto se justifica por no existir suficientes recursos para atender a todos los necesitados. Sólo es posible una cobertura parcial de la población con una intervención, o el programa continuará por etapas a través del tiempo. En otros casos, la intervención no se ha probado, de manera que es incierto que lo que no se administra sea de valor. Si se demuestra que una intervención es eficaz, puede ser difícil explicar por qué se les niega el acceso a ella a algunas personas (Patton, 2008).

Grupo de tratamiento: Grupo en un experimento cuyos miembros están expuestos a una intervención, o se les administra.

Asignación al azar

El diseño experimental implica asignar al azar participantes potenciales al programa a los grupos de tratamiento y de control, a fin de maximizar la probabilidad de que los grupos sean idénticos (no hay sesgo en la asignación a los grupos) en términos de factores que puedan influir en el programa, o en los resultados de la intervención. Estos factores podrían incluir edad, género, educación, actitud, historia de los participantes y otros factores.

En un mundo ideal, se podría decidir al azar quién recibe la intervención y quién no. Un problema del mundo real en la evaluación es el de identificar un grupo de control creíble que no reciba la intervención, y una forma de lograrlo es asignar los recursos del proyecto o el programa de forma aleatoria. Los beneficiarios del proyecto o el programa constituyen entonces una muestra al azar de la población como un todo. Esta muestra puede compararse entonces con los no beneficiarios de otra muestra extraída al azar (el grupo de control) (White, 2007).

La asignación al azar puede habilitar el uso de un diseño fuerte para medir el impacto. Puede ser también más equitativo que asignar según un método distinto; no está en juego ningún sesgo o favoritismo al asignar los grupos de tratamiento y de control con base en la probabilidad al azar.

Aunque la asignación al azar es más aplicable a las intervenciones para el desarrollo de lo que podría pensarse, no siempre constituye una alternativa. A semejanza de las pruebas médicas, a veces no es ético retener una intervención para crear un grupo de control, aunque puede no haber suficientes recursos para incluir a todos los que solicitan participación, lo que da como resultado retener el tratamiento a algunos. Un administrador de proyecto puede asignar la participación a quienes tienen mayor probabilidad de beneficiarse con la intervención, y ésta puede ser una forma de obtener el máximo beneficio de los recursos monetarios limitados del programa. Desde la perspectiva de la evaluación, sin embargo, si se asigna el programa a los mejores, los resultados estarán sesgados, porque las muestras no se crearon en forma aleatoria sin sesgos.

Variable de control: Factor ajeno que se mantiene constante para minimizar cualquier efecto posible que pudiese tener en el resultado de una intervención.

Cuando no es posible la asignación al azar, una alternativa es recopilar datos sobre factores que pueden diferir entre los dos grupos, y que parece ser probable que afecten los efectos directos. Entonces estas variables pueden incorporarse en el análisis de los datos como *variables de control.* El utilizar variables de control le permite al evaluador descartar algunas explicaciones alternativas, aun cuando no sea posible la asignación al azar.

Al seleccionar los grupos, los evaluadores deben considerar el problema del *sesgo de selección.* En éste, una diferencia entre los participantes y los no participantes puede basarse en diferencias no observables entre los grupos y no en los efectos de la intervención.

Sesgo de selección: Distorsión de la evidencia o de los datos sobre los resultados de una intervención, debida a diferencias sistemáticas en las características del subconjunto de la población que recibe la intervención y los de la misma población que no la reciben.

> El proceso de asignación al azar asegura que antes de tener lugar la intervención los grupos de tratamiento y de control sean estadísticamente equivalentes, en promedio, con respecto a todas las características. Los experimentos al azar resuelven el problema del sesgo de selección, generando un grupo de control experimental de personas que habrían participado en el programa o el tratamiento pero a quienes se les negó el acceso a éste en forma aleatoria. La asignación al azar no elimina el sesgo de selección pero equilibra los sesgos entre el grupo participante (de tratamiento) y el no participante (de control), de manera que se cancelan al calcular el estimado del impacto promedio. Las diferencias en los promedios de los efectos directos de los dos grupos después de la intervención pueden atribuirse a ésta (Banco Mundial, 2008).

El sesgo de selección puede ocurrir de dos maneras. Primero, los participantes pueden autoseleccionarse para ingresar al programa. Segundo, los administradores del programa pueden seleccionar a los participantes con mayor posibilidad de éxito. Por ejemplo, considérese el programa de los toldos para mosquitos. Si la intervención del programa fuese introducir toldos en el mercado a muy bajo costo, habría sesgo de selección, pues sólo quienes tuviesen los medios para adquirirlo y lo adquiriesen estarían en el grupo de tratamiento, y quienes no pudiesen adquirirlo o no tuviesen información sobre el programa no estarían en ese grupo. Además, sería muy difícil comparar los grupos de tratamiento y de control, pues puede no existir registro de quiénes utilizaron el toldo y quiénes no.

Para reducir la posibilidad de engañarse y pensar que algo que no es cierto sí lo es (un falso positivo), los evaluadores toman en préstamo los métodos de las ciencias sociales. Utilizando un diseño de evaluación experimental, los evaluadores hacen todo lo que pueden para controlar la implementación de un programa, una política o un proyecto y el entorno en donde se ejecuta. Cuando en la evaluación se puede controlar razonablemente todo excepto la intervención, los evaluadores pueden estar de modo razonable seguros de que las diferencias observadas son resultado de esa intervención.

Considérese una evaluación de una intervención que aplica fertilizantes para incrementar la producción del cultivo de maíz. En el proyecto se opera un invernadero en donde los aldeanos controlan la temperatura, el agua y las condiciones del suelo. Como parte del diseño de la evaluación, se crean dos zonas de cultivo separadas dentro del invernadero. Una se asigna al azar para recibir fertilizante y la otra sirve como zona de control. Ambas zonas reciben las mismas temperatura, luz solar y agua, y el maíz se planta en exactamente la misma mezcla de suelo. En la cosecha se miden las producciones de ambas zonas. Si la zona de prueba tiene mayor producción que la de control, el evaluador puede concluir que la diferencia se debió al fertilizante.

Ahora considérese qué sucede cuando se aplica la intervención en el campo en lugar de en el entorno controlado del invernadero. ¿Qué sucede si las dos zonas están muy cerca la una de la otra y el fertilizante se escapa a la zona de control? Esta zona podría moverse a una parte distinta del campo, pero entonces el suelo, la luz, la temperatura o las lluvias podrían ser ligeramente diferentes allí. Además, los dos campos pueden recibir una atención distinta. Aunque la evaluación podría todavía medir el impacto,

001El camino.indd 240 08/11/2010 03:27:50 p.m.

probablemente sería más tentativo concluir que sólo el fertilizante fue el causante de la mayor producción.

En el mundo complejo en donde tienen lugar las intervenciones para el desarrollo, se hace difícil determinar la atribución en medio de otros factores. En el caso de la agricultura, supóngase que se implementó una intervención de irrigación durante un tiempo de clima ideal y fuerte demanda de cultivos. El ingreso en la zona en donde se implementó la intervención de irrigación aumentó con respecto a los últimos años. Pero, ¿es el mayor ingreso resultado de la intervención? ¿O es causado por otros factores, como la mayor cantidad de lluvia, los buenos tiempos económicos, o un período inusual de estabilidad política? Idealmente, se tomarían los elegibles para la intervención de irrigación de agua dentro de una zona definida, y asignar al azar algunos al grupo de la intervención y otros al de control. Pero, ¿cuáles son las alternativas cuando no es posible la asignación al azar y, por tanto, el diseño experimental no constituye una opción?

En muchos casos, los diseños cuasiexperimentales se utilizan para la atribución. En estos diseños, hay una comparación pero sin la asignación al azar de los grupos. Chatterji (2007), Patton (2007) y Bamberger y White (2007) tienen excelentes artículos sobre las limitaciones de los diseños experimentales en el mundo real.

Validez interna

Al hablar acerca de eliminar otras explicaciones posibles, se trata de la *validez interna* que se refiere a la capacidad del diseño de descartar otras explicaciones para los resultados observados. Un diseño de evaluación con una fuerte validez interna aumenta la confianza de los evaluadores en su conclusión de que la intervención fue, o no, la causa de los resultados observados. Un diseño con una débil *validez interna* dificulta convencer a los demás de que la intervención fue la causante de los resultados observados. La validez interna se trae a colación aquí porque, en ausencia de la asignación al azar, la validez de los hallazgos de la evaluación puede verse comprometida en varias formas.

Validez interna: Capacidad de un diseño de descartar los demás factores o las explicaciones alternas potenciales para los resultados observados, distintos de la intervención.

Estas amenazas a la validez interna de los hallazgos de la evaluación son sólo posibles explicaciones rivales, y pueden no existir realmente, por lo que la validez interna está relacionada con el contexto. En los diseños cuasiexperimentales se deben atender las amenazas a la validez interna.

La United Kingdom Evaluation Society (2003) define la validez interna así:

> La confianza que puede tenerse en las propias conclusiones sobre lo que verdaderamente realizó la intervención. Una amenaza a la validez interna es una objeción de que el diseño de la evaluación permite el vínculo causal entre la intervención y los efectos observados sean inciertos. Puede considerarse como una cuestión de esta naturaleza: ¿No podría algo además de la intervención representar la diferencia entre la situación después de la intervención y el contrafactual?

En un texto clásico, Cook y Campbell (1979) identificaron varias amenazas comunes a la validez interna:

- Historia
- Maduración
- Pruebas repetidas
- Sesgo de selección
- Mortalidad
- Regresión al promedio
- Instrumentación.

El efecto historia. Se refiere a la posibilidad de que los eventos casos que ocurrieron durante el curso de la intervención, o entre medidas repetidas que no son parte de la intervención, pueden haber influido en los resultados. La historia será siempre una amenaza en la investigación longitudinal y es quizá la amenaza más difícil de detectar, porque el evaluador tiene que investigar eventos que ocurrieron durante la intervención que pueden haber afectado los resultados. Al observar los resultados para un individuo, los eventos históricos que afectan los resultados son bastante probables. La historia personal incluye una sucesión de eventos, algunos de pueden modificar atributos. Para un grupo de individuos, una amenaza histórica a la validez interna tiene que identificar un evento caso que simultáneamente afectó la mayoría, o por lo menos a algunos de los individuos lo suficiente para modificar de modo apreciable el atributo medido. Como anotan Brossart, Clay y Wilson (2002):

Efecto historia: Efecto de los eventos no relacionados con una intervención sobre sus resultados.

> Si todos los individuos son miembros de algún grupo, la búsqueda de este evento puede efectuarse mediante la entrevista o la observación; si los participantes son independientes, la probabilidad de que dicho evento cambie simultáneamente el atributo será pequeña a menos que el evento ocurra en un entorno común en donde se hallen ubicados todos los participantes, por ejemplo en un hospital.

Por ejemplo, considérese que durante el curso de un programa dirigido a jóvenes de alto riesgo, un joven comete un crimen atroz: la situación ocasiona un clamor por respuestas más duras a los jóvenes de alto riesgo. La situación puede alterar los tipos de personas remitidos al programa y presumiblemente afectar los resultados. Las encuestas sobre actitudes están particularmente sujetas a influencias de este tipo, porque las opiniones pueden verse fuertemente influidas por los eventos recientes y la presentación de los medios de los temas de actualidad (Office of Juvenile Justice and Delinquency Prevention, 1989).

Considérese una intervención tal como la introducción de capacitación sobre una nueva semilla, o un cultivo mejorado para agricultores de una provincia. La variable dependiente o la medida de efecto directo puede ser el mayor ingreso proveniente de los cultivos en relación con el del año anterior. Pero pudo ser que los agricultores se apegaron a sus anteriores formas de cultivo y de semillas probadas y ciertas, aun cuando su ingreso aumentara en promedio. Una investigación más profunda puede mostrar un año de clima excelente para la producción de las cosechas. El clima, y no la intervención,

fue la causa de los resultados. Los eventos externos a la intervención influyeron en los resultados. Los diseños del tipo antes y después adolecen a menudo del efecto historia.

El efecto maduración. Ocurre cuando los resultados son causados por el envejecimiento o el desarrollo. Al envejecer las personas, maduran. Al madurar, pueden sentir o responder en forma distinta a las situaciones. Los cambios que ocurren naturalmente, como resultado del paso del tiempo, incluyen los de envejecer, hacerse más inteligente y adquirir experiencia. Este efecto ocurre en individuos y en grupos. Por ejemplo, es probable que los niños sean mejores lectores en un período de dos años, aun sin capacitación adicional. Las organizaciones también se desarrollan y cambian. Estos cambios pueden formar parte de un ciclo natural de crecimiento y desarrollo y no tienen que ver con la intervención. Los diseños de tipo antes y después a menudo son débiles debido al efecto maduración.

Efecto maduración: Efecto del envejecimiento o del desarrollo en los resultados.

Puede considerarse que la ocurrencia de la maduración es de dos formas: a corto y a largo plazo. La de corto plazo se demuestra con la fatiga y el aprendizaje. La de largo plazo tiene que ver con el desarrollo psicológico, los cambios culturales y los cambios ambientales que pueden afectar las explicaciones elaboradas psicológicas. Cuando las mediciones se efectúan con una diferencia de varios meses, la maduración a largo plazo es potencialmente importante.

Por ejemplo, con una evaluación pueden investigarse los efectos de un programa de lectura de dos años en los puntajes de lectura para estudiantes de primaria. Después de dos años, las destrezas cognitivas de los estudiantes aumentarán con o sin el programa de lectura. ¿Cómo puede el evaluador estar seguro de que los mejores puntajes reflejaron el programa de lectura y no la maduración de los estudiantes?

El efecto prueba repetida. Corto plazo, ocurre cuando a los sujetos se les somete a la misma prueba antes y después de la intervención, o varias veces. Los sujetos pueden aprender a responder las preguntas, estropeando la validez de los resultados.

Considérese, por ejemplo, una intervención con la que se intenta mejorar las destrezas de los maestros en los colegios rurales. Se les somete a pruebas de desempeño al final de cada mes, y los evalúan los miembros del comité de evaluación, utilizando una lista de verificación, quienes les suministran retroalimentación mensual sobre su desempeño. Los maestros pueden mejorar en las destrezas medidas con la lista de verificación sólo por la repetición de la prueba de destreza.

Efecto prueba repetida: Efecto en los resultados causado por la exposición de los sujetos a una prueba varias veces.

Sesgo de selección. Se crea cuando un grupo de personas que se autoseleccionó para un programa se compara con otro que no lo hizo. Los dos grupos no son equivalentes,

Sesgo de selección: Distorsión de la evidencia o de los datos sobre los resultados de una intervención debida a diferencias sistemáticas en las características del subconjunto de la población que recibe la intervención, y las de los pertenecientes a la misma población que no la reciben.

porque es más probable que aquellos que se autoseleccionaron mejoren sus destrezas o modifiquen sus actitudes, aun sin la intervención que aquellos que eligieron no participar. El sesgo de selección puede estar presente aun en aquellos que eligen completar una encuesta frente a quienes no responden. El sesgo de autoselección es posible en cualquier programa en el que personas o empresas averigüen sobre el programa y se inscriban en él. Este es un riesgo del diseño cuasiexperimental.

El efecto deserción. Se refiere a quienes abandonan una intervención. Perder participantes puede crear un falso efecto de tratamiento que parece ser resultado de la intervención. De igual manera que la selección puede ser una fuente de sesgo, también puede serlo la tasa de abandono diferencial de los participantes. Aunque hay una fuerte tentación de presentar resultados sólo sobre quienes culminaron con éxito el programa, el hacerlo así producirá un grupo sesgado, porque quienes abandonan un programa probablemente tendrían un peor desempeño que quienes lo terminan. Pese a que completar un programa y obtener el efecto completo del tratamiento son insumos importantes para una evaluación, no deben obnubilar la comparación con el desempeño de un grupo de comparación.

Efecto deserción: Efecto de los que abandonan la intervención en la medida de los resultados de una intervención.

Por ejemplo, considérese un programa de educación de maestros con 400 participantes que utilizó las tasas de graduación como forma de establecer el éxito del programa. Después del programa de tres años, 25 de los participantes habían muerto por Sida. La pérdida de estos participantes redujo artificialmente la tasa de graduación, creando la impresión de que el programa tuvo menos éxito del que en realidad pudo haber tenido.

Considérese otro ejemplo, utilizando el mismo programa de educación de maestros. Supóngase que la universidad patrocinadora del programa tenía la política de que las mujeres embarazadas no podrían asistir a clase o presentar los exámenes. Entonces las mujeres embarazadas no estarían incluidas en las tasas de graduación.

El efecto regresión al promedio. Muestra una tendencia natural para los individuos que obtienen un puntaje muy alto o muy bajo para obtener un puntaje más cerca de la mitad cuando se repita la prueba. Este efecto se conoce como regresión al promedio. Si una medida no es confiable habrá alguna variación entre las medidas repetidas. Las probabilidades son que las medidas se moverán hacia la mitad y no hacia los extremos. Así, en programas que seleccionan individuos o grupos con base en los puntajes extremos, pueden esperarse cambios en el desempeño cuando el grupo "extremo" hace regresión hacia el promedio, bien se haya beneficiado o no del programa.

Regresión al promedio: Tendencia natural de los individuos con puntaje muy alto o muy bajo de obtener un puntaje más cerca al promedio, al repetirse la prueba.

Por ejemplo, supóngase que en un programa para mejorar las destrezas contables en una intervención de microcrédito se escoge a los participantes con base en su puntaje,

en una prueba de destreza aritmética, y aquellos con puntaje más alto se seleccionan para participar en el programa. Si a estos participantes se les presenta la misma prueba de aritmética después de la intervención, su puntaje bajará, porque se movería hacia el puntaje promedio.

Efecto instrumentación: Efecto en los resultados de una evaluación causado por la falta de confiabilidad del instrumento utilizado para medirlos.

El efecto instrumentación. Ocurre cuando cambia la confiabilidad del instrumento. Los cambios pueden reflejar cambios en la causa de la medición (como resultado, por ejemplo, de errores de calibración). Por ejemplo, una evaluación de un programa en el que se intenta incrementar el peso de los adultos suministrando información de nutrición puede no mostrar efectos significativos si las escalas utilizadas para medir el peso corporal no se han calibrado, o si varían en cuanto a cómo y cuándo fueron calibradas.

Diseños cuasiexperimentales

Los métodos de *diseño cuasiexperimental* pueden utilizarse para hacer una evaluación cuando no se pueden construir grupos de tratamiento y de control, utilizando la asignación al azar. Los diseños cuasiexperimentales incluyen los que utilizan grupos de comparación con características similares y los que utilizan medidas múltiples, pero no grupos de comparación; el evaluador construye grupos que son equivalentes en características importantes (género, ingreso, antecedente socioeconómico) en cuanto sea posible. Se compara entonces el desempeño de estos grupos equivalentes, algunas veces llamados *grupos de comparación* (recuadro 7.2). A veces el evaluador puede crear un grupo de comparación emparejando características básicas. Otras veces, el evaluador encontrará un grupo de comparación que no es exactamente lo mismo que el grupo que recibió la intervención, pero lo suficientemente similar para permitir alguna comparación.

Diseño cuasiexperimental: Diseño en el cual se comparan grupos con características similares, o se toman medidas múltiples del mismo grupo a través del tiempo.

No es necesario seleccionar un diseño cuasiexperimental *ex ante*, pero es preferible hacerlo. Generalmente da como resultado un diseño más fuerte si puede identificarse un grupo de comparación, y no depender sólo de medidas múltiples del mismo grupo a través del tiempo. Idealmente, los datos básicos sobre el grupo de comparación se obtienen al mismo tiempo que se obtienen para el grupo que recibe la intervención y antes de la intervención. En los diseños cuasiexperimentales *ex post*, el grupo de comparación se identifica después del inicio del programa, quizá incluso después de haber terminado. Esta es, obviamente, una situación más problemática para un evaluador.

La notación para un diseño cuasiexperimental es la misma que para el diseño experimental. La diferencia está en que hay una asignación no equivalente de sujetos a los grupos para los diseños cuasiexperimentales. En algunos casos, se coloca una *N* al principio de la línea para los grupos no equivalentes. Así, un diseño cuasiexpe-

Recuadro 7.2 *¿Funcionan los colegios administrados por la comunidad? Una evaluación del programa EDUCO de El Salvador*

Con el Programa de Colegios Administrados por la Comunidad de El Salvador (EDUCO) se aspiraba a expandir la educación rural rápidamente después de una guerra civil. La evaluación del programa se efectuó para medir los efectos de descentralizar la responsabilidad educativa a las comunidades y colegios sobre los resultados de los estudiantes.

La cuestión era si la expansión rápida a las zonas rurales se hacía a expensas de un menor aprendizaje. La evaluación comparó las medidas de resultados (resultados basados en pruebas estandarizadas de matemáticas y lenguaje) de estudiantes de tercer grado de EDUCO y colegios tradicionales, controlando las características estudiantiles y el sesgo de selección utilizando los controles estadísticos. Dado que los puntajes de las pruebas pueden no ser sensibles a corto plazo, los evaluadores también tuvieron en cuenta los días faltantes debidos a la ausencia de los maestros.

Las diferencias en resultados educativos pueden verse afectadas por varios factores. Los evaluadores necesitaban determinar si las diferencias en puntajes reflejaban diferencias en el tipo de colegio u otros factores, como características familiares (educación, tamaño de la familia, ingreso), características del estudiante (género, edad, número de hermanos), características del colegio (inscripciones, calidad de los maestros, instalaciones y finanzas) y características del maestro (antecedente educativo, años de experiencia).

Los evaluadores utilizaron datos recopilados por encuestas administradas por el Ministerio de Educación para construir un modelo con el que se midiese el impacto independiente del tipo de colegio mientras se controlaban otros factores. Utilizando un complejo modelado estadístico que controlaba los factores anteriores, los evaluadores concluyeron que los puntajes de logros de los estudiantes de EDUCO y los colegios tradicionales eran semejantes. La rápida expansión no tuvo impacto adverso en el aprendizaje. En otras palabras, los colegios administrados por la comunidad eran tan eficaces como los colegios regulares.

rimental básico en el que hay grupos de tratamiento y de comparación se escribiría en esta forma:

$$N\ O_1\ X\ O_2$$

$$N\ O_1 O_2$$

Para determinar si la diferencia se debe a una intervención, la evaluación debe mostrar que el desempeño según las medidas clave o los indicadores cambió como resultado de la intervención. Hay muchos diseños cuasiexperimentales y algunos son más sólidos que otros. Ocho de estos diseños cuasiexperimentales de evaluación se discuten a continuación:

- Diseño de tipo antes y después sin grupo de comparación.
- Diseño de comparación no equivalente previa y posterior.
- Diseño de comparación no equivalente sólo posterior.

- Diseño de comparación de series de tiempo interrumpidas.
- Diseño longitudinal.
- Diseño de panel.
- Diseño correlacional con controles estadísticos.
- Concordancia de puntajes de propensión.

Diseño de tipo antes y después sin grupo de comparación

Es una forma de medir el cambio, y se efectúa comparando las medidas después de haber empezado la intervención con las tomadas antes de empezar la intervención. Las pruebas anteriores y posteriores a la intervención constituyen medidas comunes de tipo antes y después (la medida "antes" se denomina a menudo la medida básica). La recopilación de datos básicos se llama a menudo estudio de datos básicos.

El diseño simple de tipo antes y después es un diseño cuasiexperimental débil, en el sentido de que se trata de un diseño con sólo una medida antes y después, que es insuficiente por sí misma para demostrar que la intervención sola causó el cambio. Podría ser que las personas cambiaron su comportamiento porque estuvieron siendo observadas, o que ocurrió algo más al mismo tiempo que la intervención que fue la causa real de los cambios observados.

Diseño de tipo antes y después: Diseño en el cual se comparan las medidas después de haber empezado la intervención con las tomadas antes de empezar la intervención.

Cuando existe poco cambio en el desempeño medido, los evaluadores deben titubear para concluir que la intervención no funcionó. Considérese, por ejemplo, una intervención para reducir la pobreza. La zona objetivo era tan pobre que todos eran elegibles para recibir la intervención, de manera que no hubo grupo de comparación. Al cabo de 10 años, la proporción de personas en situación de pobreza no había cambiado. El evaluador no puede concluir que no funcionó la intervención para reducir la pobreza, pues puede ser que sin la intervención (aquí nuevamente se encuentra una pregunta enmarcada sobre el contrafactual), una mayor proporción habría estado en situación de pobreza.

Usualmente se considera que las medidas de tipo antes y después no producen respuestas verosímiles a las preguntas, pues no controlan otros factores que afectan los efectos directos; con ellas se comparan sólo las condiciones antes y después. No existe comparación con y sin la intervención. Este elemento de diseño debe utilizarse, por consiguiente, sólo con otros elementos del diseño.

La notación para un diseño de tipo antes y después se escribe como sigue:

$$O_1 \; X \; O_2$$

Diseño de comparación no equivalente previa y posterior

Aunque los sujetos en los diseños cuasiexperimentales no se asignan al azar a los grupos, todavía es posible utilizar un grupo de comparación. Estos grupos de comparación

pueden denominarse entonces *grupos no equivalentes*. Todavía pueden compararse los grupos, pero los evaluadores deben considerar cuidadosamente las amenazas a la validez interna tratada anteriormente.

Para hacer más equivalentes a los grupos, los evaluadores intentan emparejarlos tan estrechamente como puedan. El emparejamiento puede hacerse, utilizando, por ejemplo, características demográficas, pruebas de destrezas, pruebas de desempeño, puntajes de juicio y otros medios. Los evaluadores pueden dar a todos los sujetos una prueba previa y luego seleccionar grupos, utilizando sus puntajes en la prueba. Por ejemplo, en una intervención para mejorar la percepción sobre los problemas de género, se administró una prueba previa que cubría conceptos y principios de percepción de género. Los puntajes se clasificaron de más alto a más bajo. De los dos puntajes más altos, uno se situó en un grupo y el otro en el segundo grupo. Entonces este procedimiento se utilizó hasta que todos los estudiantes fueron asignados a uno de los dos grupos, el cual recibió entonces capacitación adicional. En algún punto, los dos grupos fueron medidos nuevamente a fin de averiguar si el grupo que recibió la capacitación adicional tenía puntajes distintos de los del grupo que no la recibió.

La notación para una comparación emparejada, no equivalente, puede escribirse así:

$$\begin{array}{llll} \text{N} & \text{O}_1 & \text{X} & \text{O}_1 \\ \text{N} & \text{O}_2 & & \text{O}_2 \end{array}$$

Diseño de comparación no equivalente sólo posterior

El diseño de comparación no equivalente sólo posterior es un diseño más débil que el diseño de comparación no equivalente previa y posterior. A semejanza de este último, existe un grupo de comparación y, por tanto, existen datos de postintervención. Aunque esto sea preferible a no tener grupo de comparación, un problema importante es que el grupo de tratamiento o de intervención y el de comparación pueden no haber empezado en la misma ubicación. Así, mientras se sabe dónde terminaron los dos grupos, no se conoce dónde empezaron. Las diferencias entre los grupos de intervención y de comparación pueden reflejar diferencias en dónde empezaron y no el efecto de la intervención. Sin embargo, éste puede ser aún el mejor diseño que permita la situación *ex post*.

$$\begin{array}{llll} \text{N} & \text{O}_1 & \text{X} & \text{O}_2 \\ \text{N} & & & \text{O}_2 \end{array}$$

Diseño de comparación de series de tiempo interrumpidas

Con un diseño de comparación de series de tiempo interrumpidas se mide el desempeño de un grupo varias veces antes de la intervención, se administra la intervención y luego se mide el desempeño del mismo grupo varias veces después de ésta. La notación para

un diseño de comparación de series de tiempo interrumpidas dentro de un grupo sería como la siguiente:

$$O_1\ O_2\ O_3\ X\ O_4\ O_5\ O_6$$

El uso del término *interrumpidas* proviene del hecho de que mientras hay una medida continua de uno o más indicadores a través del tiempo, la medida se interrumpe con la introducción de la intervención. Las múltiples medidas antes y después de la intervención son lo que distingue a este diseño de un diseño clásico de tipo antes y después. Además, puede utilizarse con uno o más grupos de comparación.

Diseño longitudinal

Diseño longitudinal: Diseño en el se hace seguimiento a los individuos en varios puntos durante un largo período de tiempo.

En un *diseño longitudinal,* una forma de series de tiempo, se evalúa a los sujetos en varios puntos sobre un largo período de tiempo. El propósito del diseño es ver cómo cambian las cosas a través del tiempo. Por ejemplo, en el sector de los servicios de salud, puede estarse interesado en investigar las preocupaciones sobre salud a largo plazo para los hijos de madres con VIH/Sida que recibieron medicamentos para impedir la transmisión del virus. En un estudio longitudinal se haría el seguimiento de los niños a través del tiempo. Se examinarían los resultados para establecer si hubo similitudes en los problemas de salud de estos niños.

Los estudios longitudinales pueden ofrecer una riqueza de información que no se encuentra con el uso de otros diseños. Sin embargo, son costosos y de difícil ejecución y adolecen de problemas de atrición (al morir los sujetos o perder contacto con los evaluadores).

La notación para un diseño longitudinal se escribe en esta forma:

$$X\ O_1\ O_2\ O_3 \ldots$$

Diseño de panel

Diseño de panel: Diseño en el cual se hace seguimiento a un panel en varios puntos durante un largo período de tiempo.

Una forma de diseño longitudinal es el diseño de panel. En lugar de hacer seguimiento a los individuos, como en el estudio longitudinal, en un diseño de panel se hace seguimiento a la misma muestra de sujetos a través del tiempo. Por ejemplo, un programa puede estar investigando las actitudes de desplazamiento y patrones de comportamiento sobre el género a través del tiempo de los estudiantes en un colegio particular. En el diseño de panel se recopilaría información sobre actitudes de género para cada miembro de una clase desde primer grado hasta sexto.

La notación para un diseño de panel se escribe así:

$$X\ O_1\ O_2\ O_3\ \ O_4\ O_5\ O_{6\ldots}$$

Diseño correlacional con controles estadísticos

Los problemas éticos y prácticos pueden algunas veces imposibilitar el uso de un diseño experimental para la evaluación. En su nivel más simple, un *diseño correlacional* busca relaciones entre dos o más variables que no pueden manipularse, y se considera como diseño no experimental. Es con estos diseños simples de correlación que se asocia con mucha frecuencia la frase "la correlación no es igual a la causación". Pero es importante tener presente que para que exista una causa y efecto, dos variables deben estar relacionadas. Hoy se utilizan ampliamente técnicas analíticas sofisticadas en las que se utilizan formas de regresión múltiple. Las técnicas de correlación se utilizan con mucha frecuencia para crear estadísticamente grupos de comparación de manera que puedan compararse. Johnson (2002) indica que tales grupos de comparación se utilizan para responder preguntas sobre relaciones, asociaciones o causas y efectos. La coordinación de un diseño longitudinal con métodos de correlación parcial es una manera poderosa de empezar a separar inferencias causales.

Diseño correlacional: Se examina la relación entre dos o más variables que no pueden manipularse.

Por ejemplo, puede investigarse una correlación entre la cantidad de tareas en casa terminadas y el rendimiento de la prueba, si todos los miembros de la población de interés tuviesen la misma motivación. La tarea es correlacionar la parte de terminación de tareas en casa no relacionada con la motivación con la parte de los puntajes de las pruebas no relacionada con la motivación. Si se puede vincular o predecir la variable dependiente con tres variables independientes, se mantiene un vínculo causal (Garbin, 2009; psych.unl.edu/psycrs/942/q2/control.ppt).

Considérese, por ejemplo, una evaluación con la cual se busca averiguar si el porcentaje de mujeres en oficinas políticas se correlaciona con un gobierno más honesto. Podrían recopilarse los datos sobre la proporción de mujeres en oficinas políticas en distintas zonas del país y el monto de corrupción informado para determinar si las dos variables están correlacionadas. Desde luego, la evidencia correlacional sola no puede establecer la causalidad; aun si los gobiernos con más mujeres en oficinas se correlacionan con menos corrupción; sería todavía necesario descartar cualquier explicación alternativa plausible para la relación.

Dado que un diseño correlacional puede configurarse en diferentes maneras, su notación puede aparecer en varias formas. La primera notación (a) más adelante muestra un diseño con tres grupos y una observación. La segunda notación (b) muestra dos grupos, uno de los cuales recibe el tratamiento. La tercera notación (c) muestra tres tratamientos distintos (X, Y y Z), cada uno seguido de una observación.

(a) O_1
O_2
O_3

(b) O_1
$X\ O_2$

(c) $X\ O_1$
$Y\ O_2$
$Z\ O_3$

Concordancia de puntajes de propensión

Concordancia de puntajes de propensión: Diseño utilizado para medir el efecto de una intervención en los participantes del proyecto en relación con los no participantes, prediciendo la probabilidad de membresía de grupo; por ejemplo, grupo de tratamiento frente a grupo de control, con base en pronosticadores observados, usualmente obtenidos de regresión logística para crear un grupo de caso hipotético. Se utiliza también para emparejamiento, o como covariadas, solo o con otras variables de emparejamiento o covariadas.

La concordancia de puntajes de propensión se utiliza para medir el efecto de un programa en los participantes del proyecto en relación con los no participantes con características similares (White y Masset, 2005). Para emplear esta técnica, los evaluadores deben primero recopilar datos básicos. Luego, identificar características observables con probabilidad de vincularse a la pregunta de la evaluación (por ejemplo, "¿Son mayores las tasas de graduación entre las muchachas que viven cerca que entre las que viven a más de cinco kilómetros de distancia del colegio?"). Las características observables pueden incluir género, edad, condición marital, distancia del hogar al colegio, arreglos de hospedaje, número de hermanos graduados de secundaria y orden de nacimiento. Después de haberse seleccionado las variables, pueden construirse los grupos de tratamiento y comparación, emparejando cada persona del grupo de tratamiento con la del grupo de comparación que sea más similar, utilizando las características observables identificadas. El resultado son pares de individuos o de familias tan similares entre sí como sea posible, con excepción de la variable de tratamiento (White, 2007).

Existe disponibilidad de herramientas de *software* para implementar la concordancia de los puntajes de propensión, de las cuales la más utilizada es Stata (Aliendo y Kopeinig, 2005).

Diseños no experimentales

Diseño de corte transversal simple

Diseños no experimentales: Tipo de diseño de evaluación en el que no se hace intento de crear grupos de intervención y no intervención y el énfasis se hace en la descripción.

Un *diseño de corte transversal* simple muestra una instantánea en un punto en el tiempo. Esta clase de diseño se utiliza a menudo con una encuesta. Los evaluadores están interesados en respuestas de subgrupos dentro de la muestra general. Los subgrupos pueden basarse en características de subgrupos, tales como edad, género, ingreso, educación, grupo étnico o cantidad de intervención recibida. El punto de este diseño es desagregar sistemáticamente los subgrupos dentro de la muestra, de modo que los evaluadores puedan examinarlos en detalle.

Diseño de corte transversal simple: Diseño que proporciona una instantánea o imagen fija en el tiempo y en un solo punto en éste, de las características de un subconjunto de una población; a veces se le conoce como diseño *one-shot*.

Una encuesta de corte transversal selecciona una muestra de ciudadanos, beneficiarios de la intervención, o anteriores participantes en ella en un punto en el tiempo. Luego se recopilan datos sobre ellos y se informa lo que respondieron (recuadro 7.3). Algunas veces, con una pregunta se busca determinar la condición actual de personas que participaron en una intervención algunos años atrás.

Recuadro 7.3 *Uso de un diseño de corte transversal para responder preguntas descriptivas*

Al evaluar un programa diseñado para apoderar económicamente a las mujeres de modo que puedan iniciar su propio pequeño negocio, los evaluadores deseaban ver qué pensaban del programa las mujeres que lo habían recibido. Sus opiniones podrían ilustrar si lo que aprendieron en el programa para el empoderamiento económico les ayudó a iniciar un negocio viable, qué clase de negocio iniciaron y si lo que aprendieron en el programa fue útil para manejar el negocio. Con recursos limitados, los evaluadores optaron por efectuar una pequeña encuesta de recientes graduadas del programa (diseño *one-shot*). La encuesta contenía preguntas sobre las características demográficas de las participantes, de modo que pudiesen compararse las respuestas según el nivel educativo, el rango de edades y el grupo étnico.

Con un diseño de corte transversal simple se responden preguntas como las siguientes:

- ¿Tienen los participantes con diferentes niveles de educación distinta opinión sobre el valor de la capacitación?
- ¿Recibieron hombres y mujeres distintos servicios de capacitación?

Por ejemplo, una pregunta de evaluación puede enfocarse en si los subgrupos de ciudadanos o los beneficiarios de una intervención están satisfechos con los servicios que recibieron, o por qué no utilizan los servicios. Los evaluadores utilizarían este diseño para aprender cómo se diferencian los subgrupos en variables como servicios recibidos, uso de los servicios u opiniones sobre los servicios en un solo punto en el tiempo.

La notación para el diseño de corte transversal se escribe así:

$$\begin{array}{ll} X & O_1 \\ & O_2 \\ & O_3 \\ & \text{»} \end{array}$$

Diseño one-shot

En un diseño *one-shot* se observa un grupo que recibe una intervención en un solo punto en el tiempo, después de la intervención. El diseño *one-shot* se puede ver como una fotografía con la fecha impresa en ella. Este diseño puede utilizarse para responder preguntas como "¿Cuántas mujeres se capacitaron?" o "¿Cuántos participantes recibieron asesoría laboral durante un período especificado?"

Diseño *one-shot:* Diseño en el que se observa un grupo que recibe una intervención en un solo punto en el tiempo después de la intervención.

Los evaluadores utilizan diseños *one-shot* para: a) hacer preguntas a los participantes en el programa sobre si éste les gustó y cuánto, o b) determinar cómo supieron de los servicios ofrecidos. La notación para diseños *one-shot* se escribe así:

$$X \; O_1$$

Estrategias de seguimiento causal

Muchas de las estrategias para determinar si los cambios observados son causados por una intervención requieren una estrategia de recopilación estructurada y de datos cuantitativos. Para el evaluador que efectúa una evaluación rápida o una intervención muy pequeña o nueva, tales estrategias pueden no ser prácticas ni aconsejables. Aunque es mejor escoger un diseño de evaluación fuerte en cuanto ello sea posible, en situaciones como las aquí descritas, un diseño más débil, utilizando estrategias de seguimiento causal, puede ser la única opción existente.

¿Qué opciones están disponibles cuando el tamaño de la muestra es pequeño, las estrategias de recopilación de datos son en gran medida abiertas, o no es factible el análisis estadístico sofisticado? Puede presentarse uno o más de ocho argumentos lógicos para descartar hipótesis rivales. Los argumentos se refieren a la presentación de la lógica alrededor de la causalidad, y se llaman *estrategias de seguimiento causal.*

Estrategia de seguimiento causal: Tipo de diseño no experimental que consiste en argumentos para relaciones causales, basados en modelos de teoría del cambio y que descartan lógicamente explicaciones alternativas o rivales.

1. *Inferencia de lista causal*. Se sabe que un efecto directo particular es casi siempre causado por A, B, C o D. En una ocasión, no ocurrieron ni B, ni C ni D, de modo que puede estarse casi seguro de que la causa fue A. Aunque no pueda aplicarse el método aleatorio, puede recurrirse a estudios que sí lo hicieron.

   En el ejemplo del recuadro 7.1 los aldeanos sabían que los elefantes huían cuando tocaban pitos (A), golpeaban las ollas y las cacerolas (B), gritaban (C) y corrían alrededor levantando polvo (D). Si los aldeanos solo hacen (A) y los elefantes huyen, pueden estar casi seguros de que tocar pitos hace que huyan los elefantes.
2. *Inferencia de modus operandi*. Esta técnica es útil si ocurrió más de una causa posible. Supóngase que se sabe que un efecto directo es casi siempre causado por A, B, C o D, y que en esta ocasión no ocurrieron ni C ni D, estrechando la causa a A o B. Además, sólo estuvo presente la característica del patrón de eventos cadena/*modus operandi*/chisme para A. Esta inferencia se fortalece si el *modus operandi* para A es muy diferente del de B.

   Si los aldeanos aprenden de otra aldea que los elefantes no huyeron cuando los aldeanos los persiguieron y corrieron alrededor levantando polvo, pueden estar casi seguros de que la causa de la huida de los elefantes fue el toque de los pitos. Este resultado se fortalece por el hecho de que tocar pitos es muy diferente de levantar polvo.

3. *Precedencia temporal*. El efecto observado sucedió sólo después de empezar la intervención y no antes.
   Si los elefantes llegaran, los aldeanos empezaran a tocar pitos y los elefantes huyeran, entonces los aldeanos pueden pensar que hay alguna conexión entre el toque de los pitos y la huida de los elefantes. Si los aldeanos estaban tocando pitos antes de llegar los elefantes y éstos de todas formas llegaron a la aldea, entonces el toque de los pitos probablemente no causó la huida de los elefantes.
4. *Conjunción constante.* El efecto fue observado en todas las ubicaciones en que se llevó a cabo la intervención.
   Supóngase que los aldeanos se reunieron con aldeanos de toda la región, y compartieron su hipótesis de que el tocar pitos causa la huida de los elefantes. Otros aldeanos intentan esta técnica y ven que los elefantes huyen. Entonces puede estarse casi seguro de que tocar pitos hace que huyan los elefantes.
5. *Fuerza de asociación*. El cambio observado fue mucho más fuerte donde se implementó el programa que donde estaban presentes otras posibles causas.
   Si las aldeas de la región utilizan muchas técnicas distintas para espantar a los elefantes, y en las aldeas que se tocaron pitos hubo más éxito en la huida de los elefantes, puede asociarse la huida de los elefantes con el toque de los pitos.
6. *Gradiente biológico.* Cuanto más tratamiento se recibe, mayor el cambio observado.
   Supóngase que los aldeanos utilizaron más de una técnica para hacer huir a los elefantes. Cuando tocan varios pitos muy ruidosamente, los elefantes huyen. Cuando tocan un solo pito, los elefantes no huyen. Podría asociarse entonces la huida de los elefantes con el toque muy ruidoso de pitos.
7. *Coherencia.* La relación entre la intervención y el cambio observado se ajusta lógicamente a otras cosas que se conocen sobre la intervención y el efecto directo.
   Los animales peligrosos, por ejemplo, hipopótamos, cocodrilos y hienas huyen cuando los aldeanos tocan pitos. Podría concluirse lógicamente que el tocar pitos hace huir a los animales peligrosos, y se podría aplicar la estrategia a los elefantes.
8. *Analogía.* El patrón entre la intervención y los cambios observados es semejante al bien establecido patrón entre una intervención relacionada y sus efectos.
   Los aldeanos oyen una historia sobre una aldea en Suramérica que utiliza el sonido de pitos de tono alto cuando oyen un puma en la zona. Los aldeanos suramericanos creen que el ruido mantiene alejado al puma. Los aldeanos africanos pueden establecer una analogía con su problema y concluir que ruidos agudos pueden espantar a los elefantes.

En cada uno de estos casos, el principio es el mismo: el investigador descarta automáticamente explicaciones alternativas, una por una, hasta convencerse de que es más probable que los cambios observados son o no son causados (primordialmente o al menos de manera sustancial) por la intervención.

Al diseñar una estrategia de recopilación de datos, los evaluadores deberían determinar cuál de las evidencias mencionadas es factible y necesaria para recopilarla, y luego

planean cómo obtenerla. No es necesario que todas hagan contribuciones causales; los evaluadores reúnen las piezas que tienen más sentido y que juntas darán certeza suficiente sobre los hallazgos, dadas las decisiones que tendrán como base la evaluación. Pero establecer que más de una de estas estrategias de rastreo causal está presente puede fortalecer la base para inferir la causalidad.

Diseño de estudio de caso

Un *estudio de caso* es un diseño no experimental. En él no se utilizan la selección al azar ni los grupos de control y comparación. Un diseño de estudio de caso se utiliza con frecuencia cuando el evaluador desea obtener la comprensión en profundidad de un proceso, evento o situación y explicar por qué ocurrieron los resultados. Es útil cuando la pregunta trata sobre cómo funciona algo, o por qué sucede algo. Es especialmente útil cuando la intervención es innovadora o experimental o no está bien comprendida. En los casos de estudio se hace énfasis más que en las descripciones; también incluyen la interpretación de las situaciones de quienes más las conocen.

Estudio de caso: Un diseño no experimental que ofrece una descripción y una comprensión integral en profundidad de una intervención como un todo y dentro de su contexto.

Los casos de estudio se utilizan frecuentemente al evaluar intervenciones para el desarrollo. El diseño de casos de estudio es particularmente útil para describir a qué se asemeja una implementación de la intervención en el terreno, y por qué sucedieron las cosas en la forma como lo hicieron. Para examinar extremos de un programa o una intervención típica puede utilizarse un estudio de caso descriptivo.

En los casos de estudio se utilizan métodos cualitativos, cuantitativos o ambos, para recopilar los datos. Puede haber uno o varios casos de estudio en una evaluación. Su intención y su objetivo son centrarse en conocimientos profundos de los efectos de una intervención en organizaciones, comunidades, programas, ciudades o países.

Por ejemplo, para evaluar el transporte público en un país, podría simplemente hacerse un seguimiento de los indicadores clave frente a los datos básicos y los objetivos. Podría efectuarse un estudio nacional si los indicadores son el número de kilómetros cubierto por el transporte público, el número de personas que utilizan el sistema y los ingresos recibidos. Sin embargo, si hubiese otra clase de preguntas pertinentes que requieran una recopilación más profunda de datos, podría optarse por un estudio de caso.

Por ejemplo, si se solicita a un evaluador analizar un programa para mejorar el transporte hacia zonas rurales, éste podría investigar las elecciones de las personas sobre el uso del transporte público. En el diseño podría estipularse que los datos se recopilasen directamente de las personas de las zonas rurales. Serían necesarios más recursos para recopilar esos datos a una escala nacional. Es más manejable recopilarlos dentro de una geográfica definida más estrechamente (un solo caso).

De forma alternativa, los evaluadores podrían optar por varios casos de estudio, en los cuales pueden seleccionarse varias zonas rurales. Los casos pueden seleccionarse al azar o a propósito con base en algún criterio específico (el mejor caso, el caso típico, el

peor caso, incluir sólo zonas rurales aisladas cerca de grandes ciudades). Las mismas estrategias de recopilación de datos utilizadas en el estudio de caso único pueden utilizarse en varios casos de estudio.

Los casos de estudio tienen sentido para el desarrollo en donde la intención es comprender una situación específica, con el fin de formular o ajustar una política o práctica. No sólo son más prácticos los casos de estudio que los grandes estudios nacionales, sino que además proporcionan información más profunda que a menudo es útil para los encargados de tomar las decisiones (recuadro 7.4). Un estudio de caso comparativo del uso de clínicas de vacunación gratis, por ejemplo, podría ayudar a explicar por qué un enfoque es más exitoso que otro.

La notación para un diseño de estudio de caso es así:

$$O_1$$

$$O_2$$

$$O_3$$

Diseño con preguntas descriptivas

Las preguntas descriptivas son las de los tipos "¿cuántos?" o "cuánto?", y pueden solicitar percepciones u opiniones. En las preguntas descriptivas se utilizan generalmente diseños descriptivos o no experimentales. Cuando se utilizan para responder preguntas descriptivas, en estos diseños no se incluye un grupo de comparación que no recibió la intervención. Sólo se centran en quienes reciben la intervención. Algunos de los diseños utilizados para preguntas descriptivas son los mismos usados para las preguntas de causa y efecto.

Para responder preguntas descriptivas, la mayoría de los diseños comunes incluyen lo siguiente:

- De corte transversal simple.
- *One-shot.*
- De antes y después.
- De series de tiempo interrumpidas.

Recuadro 7.4 *Ejemplo de diseño de estudio de caso para preguntas descriptivas*

En un estudio en el que se investigó un programa de microcréditos en India se deseaba explorar las formas como las mujeres conceptuaban e iniciaban sus ideas de mercadeo. En el método de estudio de caso se escogieron cinco mujeres y sus proyectos y se hizo seguimiento a su progreso durante tres años.

- Longitudinales.
- De casos de estudio.

Aquí discutimos cómo funcionan algunos de estos diseños con preguntas descriptivas.

Diseño de antes y después

Los diseños de antes y después se introdujeron en una sección anterior al tratar los diseños para preguntas de causa y efecto. Estos diseños también pueden utilizarse para responder preguntas descriptivas. En un diseño de antes y después, a menudo llamado *prediseño y posdiseño*, los evaluadores hacen preguntas sobre las características del grupo antes y después de la intervención y no hay grupo de comparación (recuadro 7.5). Por ejemplo, podría preguntarse si los participantes en el programa aumentaron sus conocimientos de técnicas de cuidado de niños, y luego ponerlos a prueba en el ingreso al programa y después de terminar éste. La notación para diseños de antes y después es esta:

$$O_1 \; X \; O_2$$

Diseño de series de tiempo interrumpidas

Los diseños de series de tiempo interrumpidas se introdujeron antes al tratar los diseños para preguntas de causa y efecto. Este diseño puede utilizarse también para responder preguntas descriptivas (recuadro 7.6). En los diseños de series de tiempo interrumpidas se consideran los cambios a través del tiempo, generalmente con el fin de identificar tendencias. Cuando se utilizan para responder preguntas descriptivas, el propósito es explorar y describir cambios a través del tiempo tanto antes como después de la intervención. Así estos diseños pueden utilizarse para discernir tendencias. La notación para un diseño de series de tiempo interrumpidas es así:

$$O_1 \; O_2 \; O_3 \; X \; O_4 \; O_5 \; O_6 \ldots$$

Recuadro 7.5 *Uso de diseño de antes y después para responder preguntas descriptivas*

En un diseño de antes y después podrían observarse los salarios de una muestra de participantes en un programa de capacitación vocacional antes del programa, y dos años después del programa para responder la pregunta de cuánto aumentaron los salarios en promedio. Este diseño podría transformarse fácilmente en un diseño de antes y después de corte transversal, haciendo preguntas de subgrupos de personas con diferentes ocupaciones, con el fin de estudiar la relación de incrementos salariales a los distintos tipos de vocaciones.

Recuadro 7.6 Ejemplo de diseño de series de tiempo interrumpidas para preguntas descriptivas

Las tasas de mortalidad infantil pueden examinarse a través del tiempo, tanto antes como después de una intervención que proporciona suplementos nutritivos. Alternativamente, los cambios en las actitudes de los participantes a través del tiempo hacia mujeres empresarias pueden examinarse también tanto antes como después de la iniciación de un programa de créditos para microempresas.

Diseño longitudinal

Un *diseño longitudinal* es un tipo de diseño de series de tiempo interrumpidas en el cual medidas repetidas de la misma variable se toman de los mismos sujetos. Cuando se utiliza para preguntas descriptivas, puede utilizarse un diseño longitudinal, por ejemplo, para averiguar si los niños que asisten a un programa de enriquecimiento mantienen logros de aprendizaje a través del tiempo.

También puede utilizarse un diseño de panel para responder preguntas descriptivas. Un diseño de panel es un tipo especial de diseño longitudinal en el cual se le hace seguimiento a un grupo más pequeño de las mismas personas, en varios puntos en el tiempo y se registran sus experiencias con detalle considerable. En los diseños de panel casi siempre se utilizan técnicas cualitativas (preguntas de encuesta abiertas, entrevistas en profundidad, observación) y datos cuantitativos. Los diseños de panel pueden ofrecer una perspectiva más profunda sobre los cambios que experimentan las personas como resultado de una intervención. La notación para un diseño longitudinal es así:

Diseño longitudinal: Diseño de series de tiempo interrumpidas en el cual medidas repetidas de la misma variable se toman de los mismos sujetos.

$$X\ O_{1,}\ O_{2,}\ O_{3}\ldots$$

Diseños para preguntas normativas

La lógica para preguntas normativas es semejante a la usada para preguntas descriptivas, salvo que las preguntas normativas siempre se evalúan contra un criterio o un estándar. Los hallazgos se comparan con ese estándar, que puede incluir indicadores y objetivos. Por lo general, funcionan los mismos diseños para preguntas normativas que para preguntas descriptivas.

La auditoría del desempeño trata algunos aspectos del desempeño de una organización (Mayne, 2005, 2006). Las auditorías de desempeño pueden ser muy similares a las evaluaciones normativas. Barzelay (1997) identifica siete tipos de auditorías de desempeño, que se basan en una encuesta de los países integrantes de la OCDE. En el cuadro 7.2 se muestran cuatro de las más pertinentes.

Cuadro 7.2 *Cuatro tipos de auditoría de desempeño*

Tipo	*Unidad de análisis*	*Enfoque*
Auditoría de eficiencia	Organización o jurisdicción; elemento de proceso o programa	Identificar oportunidades para reducir el costo presupuestario de entregar productos del programa.
Auditoría de eficacia	Política, programa o elemento mayor de programa	Evaluar el impacto de políticas públicas; evaluar la eficacia de la política o programa.
Auditoría de capacidad de gestión del desempeño	Problema de administración pública de la organización o jurisdicción	Evaluar la capacidad de los sistemas y procedimientos de una jurisdicción, organización o programa para lograr los objetivos previstos.
Auditoría de información sobre el desempeño	Organización	Certificar la calidad de la información sobre el desempeño suministrada por la organización.

Fuente: adaptado de Barzelay, 1997.

El recuadro 7.7 ilustra un punto ya tratado: que cuando no es posible crear un diseño de evaluación experimental con grupos de control al azar, es aceptable recurrir a un diseño de evaluación cuasiexperimental con grupos de comparación.

Diseños para preguntas de causa y efecto

Los diseños experimentales se utilizan generalmente para tratar preguntas de causa y efecto (*ver* página 236).

Necesidad de diseños de evaluación más rigurosos

¿Qué diseños funcionan para las preguntas de causa y efecto?

Los líderes en el campo de la evaluación para el desarrollo internacional han venido debatiendo la necesidad de una evaluación más rigurosa de los programas (Davidson, 2006; Scriven, 2006; Bamberger y White, 2007). Ellos anotan que la mayoría de las evaluaciones efectuadas por agencias oficiales para el desarrollo son evaluaciones de procesos, las cuales se centran en la bondad con que esté operando un programa. El aumento en las evaluaciones participativas agregó más opiniones de los beneficiarios, pero "no produjo datos sensibles al análisis cuantitativo del impacto" (Bamberger y White, 2007, p. 58).

Los enfoques basados en resultados y el enfoque en los Objetivos de Desarrollo del Milenio (ODM) han dado como resultado mayores requerimientos para demostrar el impacto. Los requerimientos de evaluaciones de impacto, que se ocupan de los resultados causados por un proyecto, un programa o una política, han provenido de una variedad de fuentes, entre ellas las siguientes:

Recuadro 7.7 *Impacto de programas de capacitación de empleo para trabajadores despedidos*

Muchos países en desarrollo se enfrentan al problema de recapacitar trabajadores cuando las empresas de propiedad del Estado reducen su personal. Evaluar dichos programas es problemático, porque frecuentemente tienen diferentes componentes. Pueden servir a varios electorados y medir los efectos directos en diferentes formas (empleo, autoempleo, ingreso mensual, ingreso por hora). Para evaluar programas de capacitación laboral, los evaluadores plantearon las preguntas siguientes:

- ¿Tuvieron más éxito los participantes del programa que los no participantes al reingresar al mercado laboral?
- ¿Qué efectividad en costos tuvo cada programa de capacitación?

Los participantes que recibieron capacitación fueron emparejados con un grupo similar de no participantes. Se utilizaron datos administrativos, datos de encuestas y técnicas estadísticas para medir el impacto del programa de capacitación. Es posible que las personas que participaron fueran diferentes en alguna forma que representara la diferencia en cuanto a los efectos directos. Por ejemplo, quizá estuvieron más motivadas que las personas del grupo de comparación o tuvieron más experiencia laboral, lo que les pudo haber facilitado encontrar un nuevo empleo. Para fortalecer este diseño, sería necesario asignar al azar a los elegibles para participar en el programa a fin de eliminar la posibilidad de atributos diferenciales de los dos grupos.

Fuente: los autores.

- La Conferencia de Monterrey sobre Financiación para el desarrollo, de 2002, que solicitó un mayor uso de gestión basada en resultados en las agencias para el desarrollo.
- Los Acuerdos de París de 2005, que alentaron la cooperación de múltiples donantes en la promoción de las evaluaciones de impacto.
- El Poverty Action Lab (Laboratorio de Acción contra la Pobreza), que promueve el uso de diseños con asignación al azar y ofrece programas de capacitación para países en desarrollo en estos diseños.
- El Centro para el Desarrollo Mundial (CGD), que aboga fuertemente por diseños de evaluación más rigurosos, notablemente en la publicación *When Will We Ever Learn?* (CGD, 2006). El CGD expidió también una llamada para la acción por parte de agencias independientes de evaluación para asegurar más independencia y rigor en las evaluaciones para el desarrollo (Bamberger y White, 2007).

La ampliación de las fronteras del pensamiento actual sobre el desarrollo internacional y la evaluación para el desarrollo está impulsando el uso de la evaluación de impacto. Los esfuerzos del gobierno de España ilustran este nuevo pensamiento y el empuje hacia diseños de evaluación más rigurosos (recuadro 7.8). A menudo, se define

Recuadro 7.8 *El apoyo de España a los diseños de evaluación rigurosos*

España está buscando mejorar la eficacia de la ayuda con la promoción de la gestión basada en resultados en su propia agencia para el desarrollo y en países asociados. Ha implementado un nuevo programa para apoyar al Banco Mundial en la evaluación de impacto de programas innovadores a fin de mejorar los efectos directos en el desarrollo humano. El programa –el Spanish-World Bank Trust Fund for Impact Evaluation (SIEF)– es el mayor fondo fiduciario que el Banco Mundial ha establecido, centrado en la evaluación de impacto y los resultados (Banco Mundial, 2007).

"riguroso" como un experimento que incluye el uso de un grupo de control asignado al azar. Ravallion, director del departamento de investigación del Banco Mundial, anota que la asignación al azar no siempre es la respuesta que aparenta ser en un principio, en especial en el contexto del desarrollo (Ravallion, 2009). Por ejemplo, sólo pueden asignarse al azar algunas intervenciones pertinentes al desarrollo. ¿Cómo se asigna al azar la ubicación de proyectos de infraestructura y programas relacionados? Ravallion discute cómo pueden tener los experimentos al azar severos problemas de generalización, puesto que a menudo se realizan sólo para intervenciones estrechas y discretas bajo ciertas condiciones. Él describe también los efectos de las "consecuencias". Esto es, quienes seleccionan diseños al azar a menudo asumen que los no participantes no se ven afectados por el programa. Pero se sabe que los efectos de las consecuencias son penetrantes en las aplicaciones para el desarrollo. Las consecuencias aparecen por el movimiento de los sujetos hacia dentro y hacia fuera de una zona de tratamiento; es imitación a largo plazo, porque un grupo copia las prácticas de otro grupo si parecen funcionar, y los gobiernos locales intervienen y utilizan recursos liberados por las organizaciones para el desarrollo para ofrecer la misma intervención a un grupo de control designado.

Los diseños de evaluación deben ser apropiados para cada situación y cada tipo de pregunta de evaluación que se haga. Patton (2007) discute el emparejamiento de los diseños con las distintas clases de intervenciones para el desarrollo, anotando que "diferentes situaciones de impacto y diferentes propósitos de la evaluación tienen diferentes implicaciones para los métodos" (p. 1). Los evaluadores deben empezar clarificando la situación. Como anota Patton, "existen múltiples situaciones de impacto en el desarrollo y, por consiguiente, diversos enfoques en las evaluaciones de impacto" (p. 2).

Bamberger y Patton (2007) ofrecen las siguientes sugerencias para fortalecer un diseño de evaluación y tratar con las restricciones de tiempo y presupuesto:

1. Construir el diseño de evaluación sobre un modelo sólido de teoría programática. Esto puede contribuir a explicar los vínculos en la cadena causal e identificar los supuestos. Puede contribuir también a identificar factores locales económicos,

políticos, institucionales, ambientales y socioculturales para explicar las diferencias en desempeño y efectos directos.

2. Adoptar un buen diseño de métodos mixtos y combinar enfoques cualitativos y cuantitativos:

   - Utilizar datos cualitativos para triangulación a fin de proveer evidencia adicional en apoyo de los resultados cuantitativos, ayudar a enmarcar la investigación e interpretar los resultados cuantitativos.
   - Utilizar al máximo los datos secundarios, incluyendo los datos de control del proyecto.
   - En cuanto lo permitan el tiempo y el presupuesto, recopilar datos en puntos adicionales en el ciclo del proyecto.

3. Simplificar los instrumentos de recopilación de datos.
4. Utilizar creativamente los datos secundarios, usando datos de encuestas realizadas para los datos básicos o los grupos de control o comparación.
5. Considerar la posibilidad de reducir el tamaño de la muestra si ésta es deliberativa.
6. Reducir los costos de recopilación de datos, utilizando entrevistadores menos costosos, observación directa en lugar de encuestas de hogares y aprovechar los resultados de otras evaluaciones o sincronizarlas, agregando otra encuesta planeada.

Los evaluadores deben explorar las opciones para cada diseño a fin de proveer resultados más robustos (cuadro 7.3). La escogencia de un diseño correcto es crítica, porque "métodos diferentes pueden producir hallazgos completamente diferentes. El reto es calcular qué diseño y métodos son los más apropiados, productivos y útiles en una situación dada" (Patton, 2002, p. 255).

La evaluación es tanto un arte como una ciencia. Al tomar decisiones sobre diseño, los evaluadores deben tener presente que no existe el diseño perfecto; todas las evaluaciones implican compensaciones en términos de tiempo, costo y práctica. Estas compensaciones deben reconocerse proveyendo alguna evaluación de su probable impacto en los resultados y conclusiones.

Como lo implica Ravallion (2009), permitir que las preguntas de la evaluación dirijan el enfoque. Las preguntas más importantes pueden ser "¿Quién gana y quién pierde con la intervención? ¿Es semejante la intervención en el terreno a como es en el papel? ¿Cuál es la variabilidad en la implementación? ¿Sobre qué variaciones debe basarse? ¿Para quién y en qué condiciones parece funcionar la intervención?"

En suma, cada pregunta de evaluación requiere un diseño que sea experimental, cuasiexperimental o no experimental. Con un diseño experimental se intenta controlar todos los factores en el "experimento" para determinar o predecir lo que pueda ocurrir. En un diseño experimental se utiliza la asignación al azar de sujetos en por lo menos dos grupos, el de tratamiento y el de control.

El diseño cuasiexperimental es similar al experimental en el que se utilizan dos o más grupos, pero no se asignan al azar los sujetos a cada grupo. Usualmente se crea un

001El camino.indd 262 08/11/2010 03:27:51 p.m.

Cuadro 7.3 *Ventajas y desventajas de diseños experimentales comunes*

Tipo de diseño	*Ventajas*	*Desventajas*
Experimental		
Comparación	Controles a las amenazas internas a la validez	Difícil de realizar en el sector público
	Útil para establecer diferencias entre grupos; controles para historia y maduración si el grupo de comparación es concordancia estrecha	Selección y atrición que constituyen amenazas
Cuasiexperimental		
De antes y de después	Útil para suministrar un contexto para medir el cambio	Efectos de prueba, instrumentación, regresión al promedio, atrición, historia y maduración que pueden representar amenazas
No experimental		
One-shot	Útil para tratar preguntas descriptivas y normativas; los diseños múltiples *one-shot* empiezan a justificar el punto de vista	Muy débiles preguntas de causa y efecto

Fuente: los autores.

grupo de comparación para permitir el estudio de un grupo con la intervención y un grupo similar sin la intervención. Los diseños no experimentales son más descriptivos y no utilizan asignación al azar ni grupo de comparación.

Para la mayoría de las intervenciones para el desarrollo, es difícil crear un diseño que responda preguntas de causa y efecto, debido a la complejidad de la situación. Es difícil "probar" que una intervención causa el efecto observado. Los diseños de evaluación pueden ayudar a establecer el impacto de un programa hasta el punto que le permitan al evaluador un control sobre la implementación y la medición del programa. La intención es eliminar otras explicaciones posibles para los resultados observados.

Para las preguntas de causa y efecto, el evaluador debe considerar uno o más de los tipos de diseño de evaluación siguientes:

- Diseño de comparación emparejada y no equivalente.
- Diseño de series de tiempo interrumpidas.
- Diseño correlacional con controles estadísticos.
- Diseño longitudinal.
- Diseño de panel.
- Diseño de antes y después.
- Diseño de corte transversal.
- Diseño de concordancia de puntajes de propensión.
- Rastreo causal.

Las preguntas descriptivas generalmente se evalúan utilizando diseños descriptivos o no experimentales. Los diseños para preguntas descriptivas se centran sólo en responder preguntas que no tratan la atribución o emparejan el desempeño con algún criterio. Algunos de los diseños utilizados para preguntas descriptivas son los mismos que para los de preguntas de causa y efecto.

Para responder preguntas descriptivas, el evaluador debería considerar uno o más de los tipos de diseños de evaluación siguientes:

- *One-shot*.
- De corte transversal.
- De antes y después.
- De series de tiempo interrumpidas.
- Diseño longitudinal.
- Estudio de casos.

La lógica para preguntas normativas es similar a la de preguntas descriptivas, salvo que las normativas siempre se evalúan contra un estándar o un criterio.

Muchos líderes en la evaluación para el desarrollo internacional solicitan más rigor en el diseño de evaluación. A medida que los proyectos, programas y políticas se desplazan hacia la gestión basada en resultados, podría esperarse ver más diseños para evaluaciones con las que se intenta atender los asuntos de atribución.

001El camino.indd 264 08/11/2010 03:27:51 p.m.

Actividades del capítulo 7

Ejercicio de aplicación 7.1: seleccionar un diseño de evaluación

Se le ha pedido medir el impacto de construir una clínica de salud comunitaria para enseñar a los padres a tratar enfermedades familiares comunes, e identificar problemas que puedan ser graves. Los objetivos del programa son incrementar el número de padres con conocimiento básico de servicios de salud preventiva, primeros auxilios y estrategias de tratamiento temprano, y reducir el número de niños y adultos mayores cuya enfermedad se vuelva grave.

1. ¿Cuáles son los efectos directos deseados?
2. Escriba una pregunta de causa y efecto, una normativa y una descriptiva para esta evaluación.
3. ¿Qué diseño utilizaría para investigar estas preguntas? ¿Cuáles son las fortalezas y límites de este diseño? ¿Por qué es el diseño que eligió mejor que otros diseños posibles?

Ejercicio de aplicación 7.2: escoger un diseño de evaluación y una estrategia de recopilación de datos

Se le ha pedido elaborar un diseño de evaluación para un estudio de seis meses, para evaluar la eficacia de una campaña de información sobre salud preventiva en su país. La campaña consistirá en seminarios de dos días realizados por profesionales de la salud en comunidades en todo el país. El propósito de su evaluación es determinar si la campaña de información dio como resultado mejores prácticas de salud. Cuenta con un presupuesto moderado y seis asistentes de investigación para ayudarle a diseñar y efectuar la evaluación.

1. ¿Es su pregunta primaria de evaluación descriptiva, normativa o de causa y efecto? Explicar.
2. ¿Debería su estrategia de recopilación de datos ser estructurada, abierta o una combinación de ambas? ¿Por qué?
3. ¿Cómo identificaría los efectos directos más importantes que deben medirse y cómo los mediría?
4. ¿Qué diseño de evaluación utilizaría?
5. ¿Cuáles son las fortalezas y debilidades de su diseño?

001El camino.indd 266

Anexo 7.1 *Resumen de términos clave y atributos de diferentes tipos de diseño*

Términos Ipdet	*Tipos de diseño*	*Representación visual*	*Ventajas básicas*	*Desventajas básicas*
Experimental	*Los diseños experimentales se caracterizan por asignación al azar a los grupos de control e intervención*			
	Grupo de comparación *al azar*	O_1 X O_2 O_1 O_2	Fuerte validez interna, identifica cambio a través del tiempo con y sin intervención	Costos, consideraciones éticas, difícil de generalizar
	Grupo de comparación *al azar* sólo después, sin prueba antes	O_3 X O_4	Buena validez interna, ligeramente más práctico, útil para comparar resultado	No identifica cambio a través del tiempo
Cuasiexperimental	*Los diseños cuasiexperimentales son ligeramente más débiles que los experimentales con respecto a la validez, o tienen una baja validez interna. Los diseños cuasiexperimentales incluyen comparaciones pero sin asignación aleatoria*			
De antes y después sin comparación	Diseño de antes y después dentro del grupo	O_1 X O_2	Práctico; debe considerarse el contexto	Amenazas de prueba, instrumentación y regresión
Comparación no equivalente previa y posterior	Comparación de antes y después entre grupos (no equivalente)	N O_1 X O_1 N O_2 O_4	Debe considerarse el contexto; mayor confianza que con grupo de comparación	Descarta el efecto de historia, difícil de controlar para todas las variables, lo que hacen no equivalentes a los grupos
Comparación no equivalente sólo posterior	Sólo compara datos postintervención	N O_1 X O_2 N O_2	Puede ser el mejor diseño para la situación *ex post*	No se conoce dónde empezó el grupo de tratamiento o intervención
	Sólo posterior con grupo de comparación no equivalente	O_1 X O_2	Práctico; debe considerarse el contexto; control de los efectos de pruebas, instrumentación, regresión, historia	Consideraciones éticas; la selección amenaza la validez
	Sólo posterior con diferentes diseños de tratamiento	X O_1 Y O_2 Z O_3	Puede comparar intervenciones; debe tener en consideración el contexto	Permanecen muchas amenazas
Comparación de series de tiempo interrumpidas (bueno para preguntas descriptivas)	Series de tiempo (dentro del grupo)	O_1 O_2 O_3 O_4 X O_4 O_5 O_6	Amenaza de historia parcialmente controlada, maduración controlada	Amenaza de sesgo de prueba

(Continúa en la página siguiente)

08/11/2010 03:27:51 p.m.

001El camino.indd 267

(Continuación anexo 7.1)

Términos Ipdet	*Tipos de diseño*	*Representación visual*	*Ventajas básicas*	*Desventajas básicas*
	Comparación de series de tiempo entre grupos (no equivalente)		Descarta amenazas de historia, regresión al promedio reducida	Costoso, consume mucho tiempo, difícil para hacer seguimiento a las personas a través del tiempo
Longitudinal	Sin datos básicos	X O_1 O_2 O_3 ...	Hace seguimiento a los individuos a través del tiempo	Costoso, difícil para hacer seguimiento a las personas a través del tiempo
Panel	Hace seguimiento al mismo grupo en el tiempo	X O_1 O_2 O_3 O_4 O_5 ...	Información en profundidad	Puede ser costoso
Correlacional con controles estadísticos		O_1 O_2 O_3	Utiliza la estadística para establecer correlaciones entre los casos para aislar amenazas potenciales; determina relaciones importantes y variables potencialmente confusas	Requiere muestras de gran tamaño; no puede hacerse enunciado sobre causa; especulativo
Concordancia de puntajes de propensión	Comparación de participantes en la intervención con no participantes similares		Uso de programas voluntarios para la evaluación; evaluación más confiable del efecto del proyecto o programa en los participantes	Requiere grandes conjuntos de datos y capacidades de computación
No experimental	*Ideal para descripción*	*Todos los diseños no experimentales son más débiles que los diseños experimentales*		
De corte transversal	Dentro de y entre	X O_1 O_2 O_3 "	Cuadro claro de un punto en el tiempo	Sin indicación clara de lo que sucede a través del tiempo
One-shot		X O_1	Facilidad, práctica	Muchas amenazas a la validez; diseño débil
Estrategias de rastreo causal	Argumenta relaciones causales basadas en modelos de teoría del cambio y lógicos			
Casos de estudio		O_1 O_2 O_3	Información contextual en profundidad	Consume mucho tiempo; poca validez interna

08/11/2010 03:27:51 p.m.

Referencias y lecturas adicionales

Aliendo, Marco y Sabine Kopeinig. 2005. *Some Practical Guidance on the Implementation of Propensity Score Matching.* Documento de discusión 1588, IZA, Institute for the Study of Labor, Bonn. http://ftp.iza.org/dp1588.pdf.

Bamberger, Michael y Howard White. 2007. "Using Strong Evaluation Designs in Developing Countries: Experience and Challenges". *Journal of Multidisciplinary Evaluation* 4 (8): 58-73.

Banco Mundial. 1998. *Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program.* Impact Evaluation of Education Reforms Paper 8, Development Research Group, Washington, DC.

———. 2004. *PovertyNet: Evaluation Designs.* Washington, DC: Banco Mundial. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:412148~pagePK:148956~piPK:216618~theSitePK:384329,00.html.

———. 2006a. *Conducting Quality Impact Evaluations under Budget, Time y Data Constraints.* Independent Evaluation Group, Washington, DC.

———. 2006b. *Impact Evaluation: The Experience of the Independent Evaluation Group of the Banco Mundial.* Independent Evaluation Group, Washington, DC.

———. 2007. PowerNet. Spanish Impact Evaluation Fund. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:21419502~menuPK:384336~pagePK:148956~piPK:216618~theSitePK:384329,00.html.

———. 2008. *PovertyNet: Impact Evaluation, Methods and Techniques, Evaluation Designs.* Washington, DC: Banco Mundial. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:384329,00.html.

Barzelay, M. 1997. "Central Audit Institutions and Performance Auditing: A Comparative Analysis of Organizational Strategies in the OECD". *Governance: An International Journal of Policy and Administration* 103: 235-60.

Boruch, Robert.2004. "Ethics and Randomized Trials". International Program for Development Evaluation Training (IPDET); presentación, Ottawa.

Brossart, Daniel F., Daniel L. Clay y Victor L. Willson. 2002. "Methodological and Statistical Considerations for Threats to Internal Validity in Pediatric Outcome Data: Response Shift in Self-Report Outcomes". *Journal of Pediatric Psychology* 27 (1): 97-107.

Brown, Randall S. y Ellen Eliason Kisker. 1997. "Nonexperimental Designs and Program Evaluation". *Children and Youth Services Review* 19 (7): 541-66. http://www.aei.org/publications/pubID.17770/pub_detail.asp.

Campbell, D. T. y J. C. Stanley. 1963. "Experimental and Quasi-Experimental Designs for Research". En *Handbook of Research on Teaching,* ed. N. L. Cage. Chicago: Rand-McNally.

CGD (Centro para el Desarrollo Mundial). 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation.* Washington, DC.

Chatterji, M. 2007. "Grades of Evidence: Variability in Quality of Findings in Effectiveness Studies of Complex Field Interventions". *American Journal of Evaluation* 283: 239-55.

Cohen, M. 2001. "Evaluating Microfinance's Impact: Going Down Market". En *Evaluation and Poverty Reduction,* eds. O. N. Feinstein y R. Picciotto, 193-203. Nueva Brunswick, NJ: Transaction Publishers.

Cook, T. D. y D. T. Campbell. 1979. *Quasi-Experimentation: Design and Análisis for Field Settings.* Boston: Houghton Mifflin.

Davidson, E. J. 2000. "Ascertaining Causality in Theory-Based Evaluation". *New Directions for Evaluation* 87: 17-26.

———. 2006. "The RCT's Only Doctrine: Brakes on the Acquisition of Knowledge?" *Journal of Multidisciplinary Evaluation* 6: ii-v.

Garbin, Cal. 2009. *Statistical Control.* Lincoln, NE: University of Nebraska. psych.unl.edu/psycrs/942/q2/control.ppt.

Grembowski, D. 2001. *The Practice of Health Program Evaluation.* Thousand Oaks, CA: Sage Publications.

Homer-Dixon, Thomas. 1995. *Strategies for Studying Causation in Complex Ecological Political Systems.* Documento ocasional; Project on Environment, Population, and Security. American Association for the Advancement of Science, Washington, DC y University of Toronto. http://www.library.utoronto.ca/pcs/eps/method/methods1.htm.

Mayne, John. 2005. "Ensuring Quality for Evaluation: Lessons from Auditors". *Canadian Journal of Program Evaluation* 20 (1): 37-64.

———. 2006. "Audit and Evaluation in Public Management: Challenges, Reforms, and Different Roles". *Canadian Journal of Program Evaluation* 21 (1): 11-45.

Miles, M. B. y A. M. Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook.* 2ª ed. Thousand Oaks, CA: Sage Publications.

NIOSH (National Institute for Occupational Safety and Health). 1999. *A Model for Research on Training Effectiveness TIER.* Centers for Disease Control, Atlanta. http://www.cdc.gov/niosh/99-142.html.

Office of Juvenile Justice and Delinquency Prevention. 1989. *Evaluating Juvenile Justice Programs: A Design Monograph for State Planners.* Informe preparado para el U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention por Community Research Associates, Inc.

Patton, Michael Q. 2002. *Qualitative Research and Evaluation Methods.* 3ª ed. Thousand Oaks, CA: Sage Publications.

———. 2005. "The Debate about Randomized Controls in Evaluation: The Gold Standard Question". International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio.

———. 2007. *Design Options and Matching the Type of Impact Evaluation and Attribution Issue to the Nature of the Intervention: Background Discussion on Impact Evaluation for International Development Efforts.* Noviembre.

———. 2008. "The Logic of Experimental Designs and 10 Common Criticisms: The Gold Standard Debate". En *Utilization-Focused Evaluation.* 4ª ed. Thousand Oaks, CA: Sage Publications.

Powell, Keith D. 2004. "Research Methods on Psychological Science: Psychology 242". Presentación en PowerPoint, noviembre 8. Department of Psychology, University of Chicago, Chicago, IL.

Prennushi, Giovanna, Gloria Rubio y Kalanidhi Subbarao. 2002. *PRSP Sourcebook Core Techniques.* Washington, DC: Banco Mundial. http://go.worldbank.org/3I8LYLXO80.

Project STAR. 2006. *Study Designs for Program Evaluation.* Aguirre Division, JBS International, Inc. Bethesda, MD. http://www.nationalserviceresources.org/filemanager/download/performance-Measurement/Study_Designs_for_Evaluation.pdf.

Ravallion, Martin. 2009. "Should the Randomistas Rule?" *The Economists Voice* 6(2): 1-5.

Schweigert, F. J. 2006. "The Meaning of Effectiveness in Assessing Community Initiatives". *American Journal of Evaluation* 27: 416. http://aje.sagepub.com/cgi/content/abstract/27/4/416.

Scriven, Michael. 2006. "Converting Perspective to Practice". *Journal of Multidisciplinary Evaluation* 6: 8-9.

———. 2007. *Key Evaluation Checklist.* Febrero. http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf.

Stake, R. E. 1995. *The Art of Case Study Research.* Thousand Oaks, CA: Sage Publications.

Stufflebeam, Daniel L. 2004. *Evaluation Design Checklist.* Western Michigan University, Evaluation Center, Kalamazoo, MI. http://www.wmich.edu/evalctr/checklists/evaldesign.pdf.

Stufflebeam, D. L., G. F. Mdaus y T. Kellaghan, eds. 2000. *Evaluation Models: Viewpoints on Educational and Human Services Evaluation.* Boston: Kluwer.

Trochim, W. M. "The Research Methods Knowledge Base. http://www.socialsresearchmethods.net/kb.

Trochim, W. y D. Land. 1982. "Designing Designs for Research". *Researcher* 1 (1): 1-16. http://www.socialresearchmethods.net/kb/desdes.htm.

United Kingdom Evaluation Society. 2003. *Glossary of Evaluation Terms.* http://www.evaluation.org.uk/Pub_library/Glossary.htm.

Wadsworth, Y. 1997. *Everyday Evaluation on the Run.* St. Leonards, Nueva Gales del Sur, Australia: Allen and Unwin.

White, Howard. 2007. "Challenges in Evaluating Development Effectiveness". Documento de trabajo, Banco Mundial, Washington DC.

White, Howard y Edoardo Masset. 2005. "Quasi-Experimental Evaluation". Presentación en PowerPoint, febrero 16.

Yin, R. K. 1984. *Case Study Research.* Thousand Oaks, CA: Sage Publications.

Sitios web

Banco Mundial. *Doing Impact Evaluation Series.* http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,menuPK:384336~pagePK:149018~piPK:149093~theSitePK:384329,00.html#doingIE.

Campbell Collaboration. http://www.campbellcollaboration.org/.

Schweigert, F. J. 2006. "The Meaning of Effectiveness in Assessing Community Initiatives". *American Journal of Evaluation* 27: 416. http://aje.sagepub.com/cgi/content/abstract/27/4/416.

Scriven, Michael 2007. *Key Evaluation Checklist.* http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf.

Stufflebeam, Daniel L. 2004. *Evaluation Design checklist.* Western Michigan University, Evaluation Center, Kalamazoo, MI. http://www.wmich.edu/evalctr/checklists/evaldesign.pdf.

University of Northern Iowa, Department of Psychology. http://www.psych.uni.edu/psycrs/457/e2/control.ppt.

CAPÍTULO 8

Seleccionar y construir instrumentos de recolección de datos

En los capítulos anteriores se tratan las preguntas y los diseños de evaluación para emparejar las preguntas. En este capítulo se analiza la recolección de datos para responder las preguntas de evaluación.

Estrategias de recolección de datos

Los datos pueden recopilarse de muchas fuentes, que incluyen registros existentes, mediciones electromecánicas, observaciones, encuestas, grupos de discusión y juicio de expertos. Ninguna forma individual es la mejor forma. Como se ilustra en el cuadro 8.1, la decisión sobre qué método utilizar depende de:

- Lo que se necesite conocer.
- Dónde residen los datos.
- La disponibilidad de fuentes y tiempo.
- La complejidad de los datos que deben recopilarse.
- La frecuencia de la recolección.
- Las formas previstas de análisis de datos.

La elección de métodos depende en parte de la pregunta de evaluación que se va a responder, en parte, de lo bien que se comprenda la intervención y, en parte, del tiempo y los recursos disponibles. Existe una compensación entre la comprensión en profundidad proveniente, por ejemplo, de un estudio de caso (recolección intensiva de datos) y de la validez de los resultados producidos por una encuesta (recolección extensiva de datos). La recolección intensiva utiliza generalmente enfoques semiestructurados que permiten respuestas flexibles. La recolección extensiva requiere comúnmente enfoques estructurados que permiten eficiencia en muchos encuestados.

Para establecer qué tipo de datos recopilar, para la evaluación los evaluadores deben determinar lo más importante para el cliente principal. ¿El cliente está más interesado en datos numéricos sobre la condición de los colegios del país, o en un conocimiento más profundo de la situación en las zonas urbanas más pobres? Algunas veces ambas cosas son importantes, pero la disponibilidad de los recursos exige que se asigne prioridad sólo a una de ellas. Al margen de los tipos de datos utilizados, los evaluadores deben aplicar ciertas reglas (recuadro 8.1).

Enfoque estructurado

La *recopilación estructurada de datos* requiere que todos los datos se recopilen en exactamente la misma forma, lo cual es particularmente importante para evaluaciones de sitios múltiples y evaluaciones grupales. En estas evaluaciones, los evaluadores deben poder comparar hallazgos en diferentes sitios, paran sacar conclusiones sobre qué funciona dónde (recuadro 8.2). La estructura es también importante al comparar intervenciones alternativas para establecer cuál es la más efectiva en costos.

Recopilación estructurada de datos: Enfoque en el que se recopilan todos los datos en exactamente la misma forma.

Cuadro 8.1 *Muestra de tabla de decisión para método de recolección de datos*

Si necesita saber	*Tenga en cuenta*
Si los aldeanos con bajo nivel de alfabetización que participaron en el programa escriben mejor que los que tienen bajo nivel pero no participaron en él	• Recoger muestras de escritura antes y después de la intervención • Utilizar resultados de la prueba antes y después de la intervención
Si los participantes se ocupan en forma más activa en la educación de sus hijos	• Observar interacciones padres-hijo antes y después de la intervención • Preguntar a estudiantes, padres y maestros sobre participación de los padres antes y después de la intervención
Si los participantes en el programa estuvieron satisfechos con la calidad de los talleres de alfabetización y el seguimiento.	• Utilizar una entrevista estructurada de los participantes • Realizar una encuesta para determinar si los niveles de alfabetización son suficientemente altos.

Fuente: los autores.

Recuadro 8.1 *Reglas para la recolección de datos*

Al recopilar los datos, los evaluadores deberían aplicar las reglas siguientes:

- Utilizar varios métodos de recolección en cuanto sea posible
- Utilizar datos disponibles si es posible (es más rápido, menos costoso y más fácil que generar nuevos datos)
- Si se utilizan datos disponibles, averiguar cómo los recopilaron los anteriores evaluadores, cómo definieron las variables y cómo se aseguraron de la precisión de los datos. Revisar el grado de datos faltantes.
- Si deben recopilarse datos originales, establecer procedimientos y seguirlos (protocolo); mantener registros precisos de definiciones y codificaciones; efectuar pruebas previas y verificar la precisión de las codificaciones y la entrada de datos.

Fuente: los autores.

Los enfoques de recolección estructurada de datos se utilizan para recopilar datos cuantitativos, cuando el evaluador:

- Necesita tratar preguntas de extensión.
- Tiene una muestra o población grande.
- Sabe lo que debe medirse.
- Necesita mostrar resultados en forma numérica.
- Necesita efectuar comparaciones entre diferentes sitios o intervenciones.

Enfoque semiestructurado

Enfoque semiestructurado de recolección de datos: Enfoque de recolección de datos en el cual no se recopilan los datos en igual forma cada vez.

Puede ser sistemático y seguir procedimientos generales, pero los datos no se recopilan en igual forma cada vez. Las entrevistas semiestructuradas se basan, con mucha frecuencia, en un conjunto preestablecido de preguntas amplias, pero el orden para presentarlas puede depender de las circunstancias. Además, algunas respuestas se profundizan con preguntas adicionales desarrolladas durante la entrevista. Estos enfoques son más abiertos y fluidos que los estructurados, y permiten que los encuestados informen a los evaluadores lo que desean a su propio modo.

Los métodos estructurados de recolección de datos son generalmente cualitativos, y se utilizan cuando un evaluador:

- Efectúa trabajo exploratorio en un nuevo campo de desarrollo.
- Busca la comprensión de los temas o problemas.
- Desea obtener narrativas de los participantes o información más profunda.

Recuadro 8.2 *Adoptar un enfoque estructurado para evaluar una intervención agrícola*

Considérese el ejemplo de una evaluación de una intervención agrícola. Para responder una pregunta, los evaluadores deciden utilizar el contenido de humedad del suelo como medida del éxito de drenaje de la tierra. Entonces planean recopilar mediciones de contenido de humedad de varios sitios en la región, antes y después del drenaje, en el mismo período de tiempo (y con iguales condiciones de clima).

Para responder una segunda pregunta, los evaluadores utilizan una guía de entrevistas estructuradas, y preguntar a los agricultores afectados su opinión sobre efectos del proyecto. Usando las entrevistas, informan el porcentaje de encuestados con varias opiniones (tales resultados tabulados se conocen como conteo de frecuencias). Las preguntas para una entrevista estructurada deben centrarse en forma estrecha y redactarse con precisión; además, debe suministrarse un conjunto de respuestas de elección múltiple. A todos los encuestados se les debe hacer la misma pregunta en exactamente la misma forma, y solicitarles elegir entre exactamente el mismo conjunto de respuestas.

Para investigar una tercera pregunta, los evaluadores planean utilizar registros de producción de cosecha y precios a través del tiempo, antes de la intervención y después de ella, en la zona drenada y en una zona similar en la región donde no se ha efectuado drenaje a la tierra. Para investigar una cuarta pregunta, le solicitan a una muestra de 100 participantes entre 2.600 su opinión sobre el proyecto y sus efectos. Para estas entrevistas, los evaluadores planean utilizar preguntas semiestructuradas para guiar las entrevistas. Intentan indagar en las respuestas lo que sea necesario para comprender las opiniones.

Fuente: los autores.

- Desea información más profunda, rica, "detrás de bastidores".
- Busca comprender resultados de la recolección estructurada de datos que son imprevistos o no bien comprendidos, o simplemente tener ejemplos abundantes para complementar los hallazgos de una recolección estructurada de los datos.

Por ejemplo, en una evaluación de un proyecto de desarrollo dirigido por la comunidad, los evaluadores podrían elegir un enfoque semiestructurado para la recolección de datos. Dado que tales programas le dan el control de planear las decisiones a grupos locales, lo apropiado es que el evaluador utilice un enfoque semiestructurado para conocer más sobre cómo se toman las decisiones, y solicitar la opinión de los integrantes de la comunidad sobre el proceso y los efectos directos del proyecto.

Características de las buenas medidas

Los evaluadores miden creencias, actitudes, opiniones, conocimiento, destrezas, desempeño y hábitos. Al determinar cómo medir la variable de interés y recopilar datos sobre ella, los evaluadores deben tener presente cuatro temas clave:

- ¿La medida es pertinente? ¿Mide lo que importa realmente, o sólo lo que es más fácil de medir?
- ¿La medida es verosímil? ¿Suministrará información sobre la situación real?
- ¿La medida es válida? ¿Reflejará lo que el evaluador se propuso medir?
- ¿La medida es confiable? Si los datos sobre la medida se recopilan en la misma forma y de la misma fuente, utilizando cada vez las mismas reglas de decisión, ¿se obtendrán los mismos resultados?

Pertinente se refiere al grado que es aplicable lo que se mide. Los evaluadores deberían evitar medir lo que sea fácil de medir en lugar de lo que se necesita. Deben evitar también intentar medir demasiado. La matriz de diseño es una herramienta para garantizar que los datos recopilados sean pertinentes.

Verosímil es el término utilizado para indicar qué tan dignos de confianza o creíbles son los datos recopilados para la audiencia prevista del informe de evaluación. Por ejemplo, la opinión de los maestros puede no considerarse la medida más verosímil para conocer por qué son altas las tasas de abandono. Las opiniones de quienes abandonan o de sus amigos pueden considerarse como medidas más verosímiles.

Validez es el término utilizado para indicar si una medida realmente mide lo que se supone debe medir. ¿Producen las preguntas información precisa? Por ejemplo, las listas de espera tienen poca validez como medida de la demanda de ciertos programas de educación infantil, pues con frecuencia están desactualizadas y los padres colocan a sus hijos en varias listas de espera. Al inscribirse el niño, no necesariamente se borra su nombre de otras listas de espera.

Hay dos clases de validez: validez aparente y validez de contenido.

- La *validez aparente* se refiere al grado al cual el contenido de una prueba o procedimiento aparece como si midiese lo que se supone debe medir. Por ejemplo, si una evaluación mide la aptitud física, la medida de cuán rápido se corren los 100 metros puede parecer una medida válida de aptitud física.
- La *validez de contenido* se refiere al grado al cual el contenido de una prueba o procedimiento mide adecuadamente la variable de interés. Si, por ejemplo, los evaluadores intentan desarrollar una medida de la condición de salud, podrían consultar a profesionales de la salud para asegurarse de que la medida seleccionada tiene una alta validez de contenido. Una medida de la proporción real de grasa corporal de un individuo, por ejemplo, es generalmente una medida más válida de la aptitud de una persona que un reporte propio de los hábitos saludables de alimentación. Una prueba de conocimiento de hábitos saludables de alimentación puede tener mayor validez que los datos del reporte propio. Sin embargo, puede no ser una medida de alta validez porque un encues-

Validez aparente: Grado al cual el contenido de una prueba o procedimiento aparece como si midiese lo que se supone debe medir.

Validez de contenido: Grado al cual el contenido de una prueba o procedimiento mide adecuadamente la variable de interés.

Confiabilidad: Grado al cual con una medición se mide la misma cosa, en la misma forma, en pruebas repetidas.

tado no puede aplicar conocimiento de hábitos saludables de alimentación a su propia acción de comer. Los resultados podrían mostrar un cuadro más favorable de lo que realmente es.

La *confiabilidad* es el término utilizado para describir la estabilidad de la medida –el grado al cual mide la misma cosa, en la misma forma, en pruebas repetidas. Por ejemplo, las herramientas de medición para eventos deportivos deben ser muy confiables. La cinta que mide la distancia de un salto debe medir la distancia en la misma forma cada vez que se utiliza. Si es así, se considera como una medida confiable. Si no, la medida es defectuosa y podrían cuestionarse los resultados del evento.

Los pesos de los bebés, al nacer, constituyen un ejemplo de medida confiable, suponiendo que las balanzas se han calibrado en forma apropiada. Las tasas de asistencia al colegio constituyen un ejemplo de medida de baja confiabilidad, porque varían dependiendo del tiempo en que se tome la medida en el colegio.

Datos cuantitativos y cualitativos

Datos cuantitativos: Datos presentados en forma numérica.

Datos cualitativos: Datos presentados en forma no numérica.

Los datos pueden ser clasificados como cuantitativos o cualitativos. Los *cuantitativos* son datos en forma numérica y los *cualitativos* son datos en forma no numérica.

Los datos cuantitativos son datos que pueden medirse con precisión. Ejemplos de éstos son la edad, el costo, la longitud, la altura, el área, el volumen, el peso, la velocidad, el tiempo y la temperatura.

Los datos cualitativos tratan con descripciones, y pueden ser observados o autoinformados, pero no necesariamente pueden medirse con precisión. Ejemplos de éstos son los datos sobre las relaciones y el comportamiento.

Considérese una evaluación de un programa de microcréditos. Los datos cuantitativos para este programa podrían incluir el número de participantes, por género, edad, número de hijos, ingreso, inventario de producto, costo de producto y ventas. Los datos cualitativos podrían incluir descripciones de productos, relaciones familiares, conducta de los participantes, relaciones con la comunidad y sentimientos de control.

Patton (2002) identifica tres métodos de recolección de datos que pueden producir hallazgos cualitativos:

- Entrevistas en profundidad y abiertas.
- Observaciones directas (utilizando descripciones narrativas).
- Análisis de documentos escritos.

Él describe las clases de información que los evaluadores obtienen de cada uno de los tres métodos:

- Las entrevistas abiertas producen citas directas sobre experiencias, opiniones, sentimientos y conocimiento.
- Las observaciones directas pueden suministrar descripciones detalladas de actividades, comportamientos, acciones y toda la variedad de interacciones interpersonales y procesos organizativos.
- El análisis de documentos puede producir resúmenes, citas o pasajes completos de registros, memorandos y correspondencia, publicaciones e informes oficiales, diarios y respuestas abiertas por escrito a cuestionarios y encuestas.

La mayor parte de la recolección de datos cualitativos proviene de emplear tiempo en el entorno bajo estudio. El evaluador hace observaciones de primera mano de actividades e interacciones, vinculándose algunas veces a actividades como observador participante. Las notas extensas de la recolección de datos son los datos en bruto. Estos datos se organizan entonces como descripciones narrativas legibles con temas principales, categorías y ejemplos de casos ilustrativos (Patton, 2002).

La calidad de los datos cualitativos recopilados depende del evaluador. Según Patton:

> La observación sistemática y rigurosa implica más que sólo estar presente y mirar alrededor. La entrevista diestra implica más que sólo hacer preguntas. El análisis de contenido requiere considerablemente más que sólo ver lo que está allí. Para generar hallazgos cualitativos útiles y verosímiles a través de la observación, entrevistas y análisis de contenido se requieren disciplina, conocimiento, capacitación, práctica, creatividad y trabajo duro (2002, p. 5).

Patton (1987) desarrolló una lista de revisión de preguntas para ayudar a los evaluadores a decidir si los métodos cualitativos constituyen, o no, una estrategia adecuada de evaluación (recuadro 8.3). Si la respuesta a cualquier pregunta es "sí", entonces es probable que sea apropiada la recolección de por lo menos algunos datos cualitativos.

En la recolección de datos se incluyen usualmente datos cuantitativos y cualitativos, pero puede predominar uno de estos enfoques. Ambos enfoques pueden caracterizarse en las formas siguientes. Un enfoque cuantitativo:

- Es más estructurado.
- Hace énfasis en la confiabilidad.
- Es más difícil de desarrollar.
- Es más fácil de analizar.

Un enfoque cualitativo:

- Es menos estructurado.
- Es más fácil de desarrollar.
- Puede suministrar "datos ricos" (datos idiosincrásicos sobre cada unidad que se estudia).
- Demanda más intensidad laboral para recopilar y analizar los datos.
- Hace énfasis en la validez.

Recuadro 8.3 *La lista de 20 preguntas de revisión cualitativa de Patton*

1. ¿Hace énfasis el programa en los efectos directos individuales –esto es, se espera que los distintos participantes se vean afectados en formas cualitativamente diferentes? ¿Existe la necesidad o el deseo de describir y evaluar estos efectos directos de clientes individualizados?
2. Quienes toman las decisiones, ¿se interesan en esclarecer y comprender la dinámica interna –fortalezas, debilidades, y procesos generales– de los programas?
3. ¿Se necesita información detallada y profunda sobre ciertos casos de clientes o sitios de programas (por ejemplo, casos particularmente exitosos, fracasos inusuales, o casos de importancia crítica) por razones programáticas, financieras o políticas?
4. ¿Existe interés en concentrarse en la diversidad, idiosincrasia y cualidades únicas de clientes y programas individuales (en lugar de una comparación de todos los clientes o los programas contra medidas estandarizadas y uniformes)?
5. ¿Se necesita información sobre los detalles de la implementación del programa? ¿Qué experimentan los clientes del programa? ¿Qué servicios se les prestan? ¿Cómo está organizado el programa? ¿Qué hace el personal del programa? ¿Deben conocer quienes toman las decisiones lo que sucede en el programa y cómo se desarrolló?
6. ¿Se interesan el personal del programa y otras partes interesadas en la recolección de información detallada y descriptiva sobre el programa con el propósito de mejorarlo (esto es, interés en evaluación formativa)?
7. ¿Existe necesidad de información sobre los matices de la calidad del programa: información descriptiva sobre la calidad de sus actividades y resultados, y no sólo niveles, montos o cantidades de actividad y efectos directos?
8. ¿Necesita el programa un sistema de garantía de la calidad específico de cada caso?
9. ¿Están los legisladores u otras personas encargadas de tomar decisiones o proveedores de fondos interesados en que los evaluadores realicen visitas a los sitios del programa, de forma que las evaluaciones los ojos y oídos puedan ser sustitutos para ellos por estar demasiado ocupados en realizar las visitas, y para quienes carecen de las destrezas de observar y escuchar con que cuentan los evaluadores capacitados? ¿Se requiere un seguimiento legislativo de cada caso?
10. ¿Constituye una preocupación la importunidad de la evaluación? ¿Será inoportuna la administración de instrumentos estandarizados de medidas (cuestionarios y pruebas), en contraste con la recolección de datos mediante observación natural y entrevistas abiertas? ¿Generará la recolección de datos cualitativos menos reactividad en los participantes que la recolección de datos cuantitativos? ¿Hay necesidad de observaciones discretas?
11. ¿Hay necesidad y deseo de personalizar el proceso de evaluación, utilizando métodos investigativos con énfasis en contacto personal con el programa, métodos que puedan percibirse como "humanistas" y personales porque no rotulan ni numeran a los participantes, y éstos los consideran naturales, informales y comprensibles?
12. ¿Es apropiado un enfoque de evaluación sensible, es decir, un enfoque especialmente sensible a recopilar datos descriptivos y presentar información en términos de las distintas perspectivas de las partes interesadas, basado en el contacto personal directo con esas distintas partes interesadas?
13. ¿Son las metas del programa vagas, generales, no específicas, y que indican la posible ventaja de un enfoque de evaluación libre de metas que sólo recopilaría información sobre qué efectos está causando el programa realmente, en lugar de medir el logro de metas?

(Continúa en la página siguiente)

(Continuación recuadro 8.3)

14. ¿Existe la posibilidad de que el programa esté afectando a los clientes o participantes en formas imprevistas, y/o causando efectos colaterales inesperados, indicando la necesidad de un método de investigación que pueda descubrir efectos más allá de los declarados formalmente como deseables por el personal del programa? (nuevamente, una indicación de la necesidad de alguna forma de evaluación libre de metas)
15. ¿Existe carencia de instrumentación cuantitativa demostrada para efectos directos importantes del programa? ¿Está el estado de la ciencia de la medición de manera tal que no existe disponibilidad de un instrumento estandarizado válido, confiable y verosímil, o que pueda desarrollarse con agilidad para medir cuantitativamente los efectos directos particulares del programa para los que se requieren los datos?
16. ¿Es exploratoria la evaluación? ¿Se encuentra el programa en una etapa de preevaluación donde las metas y contenido del programa se están desarrollando aún?
17. ¿Se requiere una evaluación de la evaluabilidad para establecer un diseño de evaluación final?
18. ¿Existe necesidad de agregar profundidad, detalle y significado a los hallazgos estadísticos o generalizaciones de encuestas?
19. ¿Se ha hecho la recolección de datos de evaluación cuantitativos tan rutinaria que nadie le presta ya mucha atención a los resultados, sugiriendo una posible necesidad de romper la vieja rutina y utilizar nuevos métodos para generar nuevas percepciones sobre el programa?
20. ¿Existe necesidad de desarrollar una teoría programática, fundada en observaciones de las actividades e impactos del programa y la relación entre tratamiento y efectos directos?

Fuente: Patton, 1987.

El enfoque utilizado dependerá de las metas de la evaluación (cuadro 8.2). En la práctica, los datos cuantitativos y cualitativos están relacionados. De acuerdo con Trochim (2006), "Todos los datos cuantitativos se basan en juicios cualitativos y todos los datos cualitativos pueden describirse y manipularse numéricamente". En realidad, el análisis computarizado (análisis de contenido) de los documentos escritos se centra en la frecuencia de varias palabras, tipos de palabras y estructuras, convirtiendo los datos cualitativos en cuantitativos.

Método inoportuno: Método utilizado para observar participantes con su conocimiento.

Los datos pueden recopilarse de manera inoportuna o discreta. Los *métodos inoportunos* son observaciones efectuadas con el conocimiento del participante. Tales métodos se utilizan para medir percepciones, opiniones y actitudes por medio de entrevistas, encuestas y grupos de discusión. Las observaciones realizadas con el conocimiento de quienes son observados son también inoportunas.

Si en una evaluación se utilizan cuestionarios para recopilar datos, los sujetos saben que se les está estudiando, lo que puede producir resultados artificiales. Según Patton (1987, p. 33), "El instrumento mismo puede crear una reacción que, debido a su intrusión e interferencia con la operación normal del programa y el funcionamiento del cliente, omite reflejar con precisión lo que se ha logrado con el programa".

Método discreto: Utilizado para observar participantes sin su conocimiento.

Quienes están siendo estudiados pueden modificar su comportamiento o sus respuestas. Por ejemplo, un maestro cuya lección está siendo observada por el administrador principal del sistema escolar, puede enseñar de manera muy distinta a cuando no lo observa el administrador.

Los *métodos discretos* son observaciones realizadas sin el conocimiento del participante. Ejemplos de métodos discretos son los de utilizar datos de documentos o archivos y observar a los participantes sin su conocimiento (recuadro 8.4).

Herramientas de recolección de datos

La técnica elegida de recolección de datos dependerá de la situación. Sin importar qué método se elija para recopilar datos de personas, toda la información acumulada está potencialmente sujeta al *sesgo.* Esto significa que cuando se solicita información a una persona, ya sea sobre ella misma o sobre otros, la persona puede o no responder con la verdad, intencional o no intencionalmente, y distorsionar la verdad por no recordar con precisión o por temor a las consecuencias de dar una respuesta verdadera. Puede estar también ofuscada o incómoda para admitir ciertas cosas que considere no aceptables socialmente. Todos los datos autoinformados están sujetos a este problema.

Sesgo: La distorsión intencional o no de los datos en términos de su recolección, análisis o presentación.

Los encuestados pueden sentirse ofuscados para responder verdaderamente a preguntas sobre, por ejemplo, el uso de protección durante la relación sexual o la fecha de su última visita al médico. Pueden describir lo que piensan que el evaluador desea oír en lugar de la verdad.

También puede existir el sesgo de selección, el cual es el hecho de que las personas que eligen participar en un programa sean diferentes de quienes eligen no participar (este tema se trató en el capítulo 5). Con mucha frecuencia esto es un problema en las encuestas, entrevistas y los grupos de discusión. Aquellos que de manera voluntaria participan pueden ser sistemáticamente diferentes de quienes no lo hacen.

Cuadro 8.2 *Cuándo utilizar un enfoque cuantitativo y uno cualitativo*

Si usted	*Utilice el enfoque*
Desea efectuar análisis estadístico	Cuantitativo
Desea ser preciso	
Sabe lo que desea medir	
Desea cubrir un grupo grande	
Desea información narrativa o profunda	Cualitativo
No está seguro de lo que puede medir	
No necesita cuantificar resultados	

Fuente: los autores.

Recuadro 8.4 Medir la popularidad de exposiciones de arte, utilizando métodos inoportunos y discretos

La directora de un museo deseaba saber cuál de las exposiciones del museo era más popular. Para averiguarlo, podía entrevistar personas que salían del museo y preguntarles qué habían visto. Sin embargo, tal autoinformación puede ser engañosa porque los entrevistados pueden indicar las exposiciones más famosas sólo porque son las únicas cuyos nombres pueden recordar, o distorsionar intencionalmente su respuesta por el deseo de que el entrevistador las perciba en forma favorable. Para evitar estos problemas, en un estudio legendario se midió la popularidad de las exposiciones de un museo, midiendo el uso de los baldosines del piso frente a las exposiciones. Con este método discreto se evitó la distorsión introducida por el autoinforme.

Usualmente, se utiliza más de un solo enfoque de recolección de datos para responder diferentes preguntas de evaluación, o para proveer varias fuentes de datos como respuesta a una sola pregunta de evaluación. Por ejemplo, en la evaluación pueden recopilarse datos disponibles de los registros de producción de cosechas de los agricultores, entrevistar a los compradores de los productos de las cosechas, o efectuar estudios de casos para desarrollar temas para un cuestionario, o encontrar explicación de los resultados de una encuesta.

La recolección de la misma información, utilizando diferentes métodos, para incrementar la precisión de los datos se denomina *triangulación de métodos.* Los evaluadores la utilizan para fortalecer sus hallazgos. Cuanta mayor información recopilada soporte un hallazgo, utilizando distintos métodos, más fuerte será la evidencia.

La triangulación de métodos no es el único tipo de triangulación. Denzin (1978) identifica varias clases, incluyendo la *triangulación de fuentes* y *la triangulación del evaluador,* en la que participan varios evaluadores en una investigación.

Triangulación de métodos: recolección de la misma información utilizando diferentes métodos con el fin de incrementar la precisión de los datos.

Triangulación de fuentes: recolección de la misma información en varias fuentes con el fin de incrementar la precisión de los datos.

Triangulación del evaluador: recolección de la misma información de más de un evaluador con el fin de incrementar la precisión de una evaluación.

Los sujetos de un estudio no son las únicas personas que pueden afectar los resultados de una evaluación. Los evaluadores y el entorno de la evaluación pueden también tener efectos. Por ejemplo, las mujeres pueden responder en forma distinta a un entrevistador masculino que a uno femenino, y así mismo responder de manera distinta si son entrevistadas a solas o con su esposo.

En las secciones siguientes se describen nueve herramientas de recolección de datos, a saber:

- Herramienta 1: recolección participativa de datos.
- Herramienta 2: análisis de registros y análisis secundario.

- Herramienta 3: observación.
- Herramienta 4: encuestas y entrevistas.
- Herramienta 5: grupos de discusión.
- Herramienta 6: diarios, bitácoras y listas de revisión autoinformadas.
- Herramienta 7: juicio de expertos.
- Herramienta 8: técnica Delphi.
- Herramienta 9: otras herramientas de medición.

Herramienta 1: recolección participativa de datos

Los enfoques de recolección participativa de datos incluyen intensamente grupos o comunidades en la recolección de los datos. Ejemplos de esta herramienta son las reuniones comunitarias, la cartografía y las caminatas transversales.

Reuniones comunitarias

Uno de los métodos más comunes de recolección participativa de datos es el de reuniones comunitarias. Éstas permiten a los miembros de la comunidad hacer preguntas, comentarios y discutir temas de importancia para la comunidad.

Para que las reuniones produzcan datos utilizables deben organizarse bien. El evaluador y las partes interesadas deben estar de acuerdo sobre el propósito de la reunión, y comprometerse a estar presentes en ella. Antes de la reunión, el evaluador debe anunciar las reglas de terreno. Los asuntos que deben tenerse en consideración son cómo identificar los interlocutores y el formato de las preguntas y respuestas. Si la población está lo suficientemente alfabetizada, deben ponerse por escrito las reglas de terreno y la disponibilidad de los recién llegados.

La reunión comunitaria debe publicitarse ampliamente, mediante volantes, anuncios de prensa y radiales. Los integrantes de la comunidad pueden ser también responsables de difundir la voz. Los evaluadores no deberían depender primordialmente de los funcionarios locales para invitar a las personas, pues podrían invitar una selección sesgada que refleje sus propios puntos de vista. Para la reunión, la ubicación debe elegirse con la idea de estimular la participación comunitaria, y cumplir al mismo tiempo con las necesidades de comodidad, acceso y seguridad de los participantes (Minnesota Department of Health, 2007).

En las reuniones comunitarias existen varias ventajas:

- Pueden suscitar la credibilidad del proceso al mejorar la apertura y la inclusión.
- Son económicas y relativamente fáciles de arreglar.
- Permiten una participación amplia.
- Su entorno más relajado puede aumentar la participación comunitaria.
- Pueden elevar el nivel de percepción y comprensión de la evaluación y construir apoyo para él.

Cuadro 8.3 *Ventajas y problemas de las reuniones comunitarias*

Ventajas	Permiten a los miembros de la comunidad conocer la intervención y discutir los problemas
	Pueden aumentar la percepción y la credibilidad
	No son costosas
	Pueden aumentar el conocimiento del evaluador de temas importantes
Problemas	Pueden no representar con precisión a la comunidad, debido a problemas de género y poder

Fuente: los autores.

- Pueden incrementar el conocimiento del evaluador sobre importantes temas del programa.
- Pueden revelar problemas que ameriten investigación adicional.

Las reuniones comunitarias tienen también deficiencias (cuadro 8.3). Por ejemplo, los integrantes de la comunidad que eligen los participantes pueden no ser representativos de la comunidad y algunas personas con buenas ideas o un conocimiento claro de los temas pueden preferir no asistir o hablar en tales eventos. Aquellos que asisten pueden ser quienes se sienten más fuertemente relacionados (positiva o negativamente) con el programa. Quienes hablan pueden ser los que tienen posiciones de poder en la comunidad. Dependiendo del tiempo en que se efectúe la reunión, los problemas de género pueden afectar la asistencia. Es probable que la asistencia y la participación de la mujer estén relacionadas con la cultura. Debido a estos problemas, las reuniones comunitarias nunca deben ser el método primario de recolección de datos para la evaluación.

Cartografía

El dibujo o el uso de mapas existentes –proceso llamado cartografía cuando se aplica a la recolección de datos– puede utilizarse para generar discusiones sobre prioridades de desarrollo local y el grado al cual se han cumplido. Puede utilizarse para verificar fuentes secundarias de información. La cartografía puede también capturar cambios percibidos a través del tiempo (antes y después de una intervención, por ejemplo). Mientras el proceso de cartografía se aplica a menudo a la planeación de intervenciones, puede utilizarse igualmente en evaluaciones.

La cartografía puede efectuarse con individuos o grupos. Como herramienta de grupo es útil para evaluaciones participativas que incluyen a las partes interesadas, ya que provee una forma de que ellos trabajen juntos. Al mismo tiempo, la cartografía puede incrementar el conocimiento que tenga cada uno de la comunidad. Esto es especialmente importante cuando las personas tienen distinto conocimiento de la comunidad

Cartografía: Dibujo o uso de mapas existentes como parte de la recolección de datos.

basado en su condición y experiencia. La cartografía es así mismo una útil herramienta de recolección de datos para utilizarse con grupos con problemas de alfabetización.

Los tipos de cartografía incluyen los siguientes:

- De recursos
- Histórico
- Social
- De salud
- De riqueza
- De uso de tierras
- Demográfico

La cartografía de recursos se utiliza a menudo para recopilar y trazar información sobre la distribución, el acceso y el uso de los recursos dentro de una comunidad. Puede utilizarse antes y después de una intervención para identificar cambios.

Los mapas pueden desarrollarse en muchas formas. Por ejemplo, para un programa de suministro de agua y saneamiento, los evaluadores pueden desarrollar dibujos sencillos con papel y lápiz que muestren las fuentes de agua conocidas, los sitios de los retretes, las características ecológicas primarias y los patrones de asentamiento. Los evaluadores pueden entonces solicitar a los aldeanos utilizar otros materiales que ellos suministren, como hilo, paja, cinta y cuerda, para demarcar qué familias utilizan cuáles fuentes de agua para beber o los sitios de los retretes. Los lápices de colores, la tinta o los crayones pueden utilizarse para dibujar puntos o indicar la distribución de grupos especiales, por ejemplo, mujeres pobres, personas ricas o líderes (Narayan, 1996).

Ejemplos de uso de cartografía en forma participativa incluyen los siguientes:

- Solicitar a los estudiantes de un colegio que mapeen su comunidad, dibujando sus casas y las de sus vecinos.
- Solicitar a un grupo de hombres y mujeres utilizar materiales disponibles localmente, como barro, hierba, arcilla, piedras y palos, para hacer un modelo de su comunidad y marcar los recursos de agua (u otros recursos).
- Iniciar una discusión e incluir personas en el proceso de desarrollar un mapa de la comunidad durante una reunión grupal (Narayan, 1996).

La *cartografía social* implica "dibujar" una imagen conceptual de los elementos que conforman una comunidad, incluyendo sus recursos y activos, y cómo interactúan entre sí. La cartografía social puede utilizarse para presentar información sobre el diseño de la aldea, su infraestructura, demografía, grupos etnolingüísticos, instalaciones comunitarias, patrones de salud, riqueza y otros temas comunitarios. En este enfoque se reúne a los integrantes de la comunidad para comprenderlos mejor, y cómo se ajusta o no la intervención dentro de ella. Puede utilizarse como parte de cualquier enfoque si es apropiado para las preguntas de la evaluación. También puede utilizarse para comparaciones antes de la intervención y después de ésta.

Cartografía social: El "dibujo" de una imagen conceptual de los elementos que conforman una comunidad, incluyendo sus recursos y activos, y cómo interactúan.

Otras herramientas de cartografía incluyen la fotografía aérea, incluyendo Google Earth (recuadro 8.5), planimetría, mapas preparados por cartógrafos profesionales y

Recuadro 8.5 *Uso de Google Earth para la cartografía*

Con Google Earth, un programa gratis que puede descargarse en Internet, ha mapeado todo el planeta, pegando imágenes obtenidas por satélite, fotografía aérea y sistemas de información geográfica en una imagen de un globo tridimensional. Están disponibles muchas grandes ciudades en resolución lo suficientemente alta para ver edificios y casas individuales y aun autos. El grado de resolución se basa en parte en los puntos de interés.

Imagen de Google Earth de Dacca, Bangladesh

Google Earth puede ser útil para recopilar datos básicos y de tendencias. Los evaluadores pueden ubicar una zona con Google Earth, guardar la imagen e imprimirla. Los datos recopilados en fechas posteriores pueden compararse entonces con la imagen básica, a fin de mostrar los cambios a través del tiempo en, por ejemplo, vías o la provisión de electricidad a comunidades rurales.

mapas preparados por trabajadores de campo que caminan a través de una comunidad y buscan asistencia de personas clave locales. La información obtenida en esta forma puede utilizarse en actividades de planeación de varias comunidades (Narayan, 1996).

Caminatas transversales

Caminar a través de una zona que se está estudiando es muy distinto a observar una imagen aérea. En lugar de utilizar una cámara, el(la) caminante utiliza sus destrezas de observación para desarrollar un conocimiento de la zona. Seguir rutas sucias que los integrantes de la comunidad toman, revela por qué muchas mujeres van por agua sólo una vez al día o cambian de ruta al empezar las lluvias. Una caminata puede llevar al conocimiento de las divisiones sociales de la comunidad, por ejemplo divisiones de poder y de casta. Puede también ayudar a identificar la organización espacial, los estilos arquitectónicos, el uso del espacio, el saneamiento ambiental, la sobre o subutilización de instalaciones y actividades alrededor de las instalaciones de agua y saneamiento (Narayan, 1996).

Una *caminata transversal* es un tipo de caminata que un evaluador podría hacer alrededor de una comunidad, a fin de obtener una observación representativa de su gente,

Caminata transversal: Caminata para observar las personas, alrededores y recursos de una comunidad.

alrededores y recursos. Las caminatas transversales son una clase de herramienta espacial de recolección de datos. Una caminata transversal puede hacerse en tan corto tiempo como una hora o tan largo como un día.

La caminata transversal se planea dibujando una "línea transversal" a través de un mapa de una comunidad. La línea debe cruzar o cortar transversalmente todas las zonas de la comunidad para proporcionar una vista representativa de ésta.

El evaluador, acompañado por varios miembros de la comunidad, camina a lo largo de la zona representada por la línea transversal en el mapa. Él o ella habla con los miembros de la comunidad mientras observa condiciones, personas, problemas y oportunidades (Academy for Educational Development, 2002).

Los siguientes constituyen ejemplos de aspectos de una comunidad que pueden observarse durante una caminata transversal:

- Condiciones de vivienda.
- Presencia de niños callejeros.
- Naturaleza del trabajo infantil.
- Comercio callejero informal y prostitución.
- Disponibilidad de transporte público.
- Tipos de ONG y religiosas.
- Tipos de almacenes.
- Tipos de alimentos vendidos en mercados abiertos.
- Condiciones sanitarias.
- Interacciones entre hombres y mujeres.
- Presencia de instalaciones de salud.
- Instalaciones comunitarias (Academy for Educational Development, 2002).

Las caminatas transversales le ofrecen a un evaluador una visión del "cuadro general" de la comunidad, y le ayudan a identificar problemas que requieren investigación adicional.

Herramienta 2: análisis de registros y datos secundarios

Algunas veces los datos que se han recopilado pueden utilizarse para responder preguntas de la evaluación. Al utilizar conjuntos de datos recopilados por otros es necesario comprender cómo se recopilaron, cómo se definieron las variables y cómo se codificaron y limpiaron los datos, incluyendo el tratamiento que se le dio a la falta de datos, la omisión de respuestas y las bajas tasas de respuesta. Las fuentes de datos secundarios incluyen no sólo conjuntos de datos de actividades anteriores de recolección de datos, sino también artículos de periódicos, presentaciones de televisión, páginas web, *blogs* y grupos de discusión en Internet.

Uso de registros

Las agencias oficiales, clínicas, colegios, asociaciones y organizaciones para el desarrollo son sólo unas pocas de las organizaciones que producen registros. Éstos pueden constituir un fundamento para las evaluaciones.

Los registros organizativos constituyen una fuente común de información para la evaluación. La mayoría de las organizaciones ya han recopilado y organizado datos de clientes y comunidades. Pueden también haberlos resumido y publicado en informes de gerencia, documentos presupuestarios, informes al público o agencias de financiación, o informes de evaluación o de seguimiento.

McCaston (2005) presenta una muestra de tipos de datos e información secundaria asociados comúnmente con el análisis de la pobreza:

- Demográficos (población, tasa de crecimiento de la población, rurales/urbanos, género, grupos étnicos, tendencias migratorias).
- Discriminación (por género, grupo étnico, edad).
- Igualdad de géneros (por edad, grupo étnico).
- Política ambiental.
- Entorno económico (crecimiento, razón de endeudamiento, términos de comercio).
- Niveles de pobreza (relativa y absoluta).
- Empleo y salarios (formales e informales; variables de acceso).
- Sistemas de medios de vida (rurales, urbanos, en la parcela, fuera de ella, informales).
- Variables y prácticas agrícolas (lluvias, cosechas, tipos y usos del suelo, irrigación).
- Salud (desnutrición, mortalidad infantil, tasa de vacunación, tasa de fecundidad, tasa de prevalencia de anticonceptivos).
- Servicios de salud (número, nivel, razón de instalaciones a población, acceso por género, grupo étnico).
- Educación (tasa de alfabetización de adultos, tasa de inscripción escolar, tasas de abandono escolar, razón de hombres a mujeres, razón étnica).
- Colegios (número y nivel, razón de colegios a población, acceso por géneros, grupos étnicos).
- Infraestructura (vías, electricidad, telecomunicaciones, agua, saneamiento).
- Condición y problemas ambientales.
- Prácticas culturales perjudiciales.

La información anterior puede encontrarse en:

- Archivos y registros.
- Bases de datos de computadores.
- Informes industriales.
- Informes oficiales.

- Otros informes, incluso evaluaciones anteriores.
- Censos y datos de encuestas de hogares.
- Listas de correo y grupos de discusión electrónicos.
- Documentos (presupuestos, políticas y procedimientos, gráficas organizativas, mapas).
- Periódicos e informes televisivos.

Para extraer información de documentos basados en el papel, los evaluadores desarrollan un instrumento de recolección de datos que especifica con exactitud qué datos recopilar del archivo o del registro y cómo codificarlo. El objetivo es desarrollar un instrumento simple y claro. Una vez desarrollado éste, debe someterse a una prueba preliminar.

Por ejemplo, considérese una evaluación de si las enfermeras de atención crítica capacitadas en un programa patrocinado por el gobierno son más eficaces que otras enfermeras de atención crítica. Podría utilizarse un instrumento de recolección de datos para recopilar en forma sistemática datos pertinentes en sus archivos (gráfico 8.1). Los evaluadores podrían seleccionar una muestra de clínicas de atención crítica en la cual se capacitó a una o más enfermeras mediante el programa del gobierno, y revisar los registros de todas las enfermeras. Estos registros incluyen sus antecedentes educativos, el tiempo que han estado ejerciendo la enfermería y sus clasificaciones de desempeño.

Al trabajar con documentos que describen actividades o prácticas corrientes, el evaluador debería verificar que los documentos reflejen con precisión las prácticas. Las observaciones (si el programa está en curso todavía) y entrevistas pueden contribuir a esto. Por ejemplo, al utilizar documentos en donde se describe un programa de capacitación, el evaluador podría verificar si las clases se dictaron realmente cinco días a la semana, hubo disponibilidad de materiales y las participantes fueron tan diversas como indicaban los documentos. Esta verificación podría surgir de unas pocas entrevistas a personas con conocimientos del programa, su historia e implementación. Dependiendo del caso, podría ser importante buscar personas con diferentes papeles, incluso personal de presupuesto y clientes, cuyas perspectivas podrían aclarar la información obtenida y ofrecer perspectivas sobre la historia no escrita. Cuando no sean factibles ni observaciones ni entrevistas, el evaluador podría verificar información, identificando y revisando otros documentos de soporte que suministren información similar.

Es usualmente necesario leer y analizar documentos oficiales además de archivos. Por ejemplo, al evaluar un programa para mejorar la respuesta de funcionarios del gobierno a las llamadas telefónicas de los ciudadanos, los evaluadores examinarían documentos oficiales para determinar lo siguiente:

- ¿Cuándo fue autorizado el programa?
- ¿Qué metas y objetivos se presentaron para el programa en el documento que autorizó?
- ¿Cuántos funcionarios oficiales debían participar?
- ¿Qué agencias o departamentos debían participar?

Gráfico 8.1 *Muestra de instrumento de recolección de datos*

Fecha _____

ID N°: _____

1. Nivel educativo más alto terminado: _____
2. ¿Enfermera registrada? Sí _____ No _____
3. ¿Completó la capacitación oficial? Sí _____ No _____
4. Si sí, ¿año de capacitación terminado? _____
5. ¿Cuántos años de enfermera en esta clínica? _____
6. ¿Cuántos años de enfermera en otra parte? _____
7. Clasificación de desempeño en los últimos cinco años:
   Año: _____ Clasificación: _____
   Año: _____ Clasificación: _____
   Año: _____ Clasificación: _____
   Año: _____ Clasificación: _____
   Año: _____ Clasificación: _____
8. Premio por desempeño recibido en los últimos cinco años:
   ____Año ____No
   Si sí, número de premios recibidos en los pasados 5 años: _____
9. Género: ____Masculino ____Femenino

10. Comentarios: _____
    _____
    _____

Fuente: los autores.

- ¿En cuánto tiempo debería implementarse el programa?
- ¿Qué medidas de desempeño o indicadores fueron incluidos o especificados?
- ¿Cuánto fue el presupuesto para el programa?
- ¿Qué actividades debían llevarse a cabo?

Uso de conjuntos de datos existentes

Los conjuntos de datos electrónicos recopilados por una organización con mucha frecuencia son obtenidos y analizados nuevamente por otra para responder nuevas preguntas. Esta clase de análisis se denomina análisis de datos secundarios. Los evaluadores utilizan a menudo grandes bases de datos computarizadas que contienen, por ejemplo, datos de encuestas de hogares o de préstamos efectuados

Análisis de datos secundarios: Análisis de datos obtenidos de otra fuente en lugar de ser recopilados por los evaluadores.

por un intermediario financiero a pequeñas y medianas empresas. Al realizar análisis de datos secundarios, los evaluadores deben mantenerse concentrados en el propósito y diseño de la evaluación.

Temas clave para tener en consideración al determinar si obtener y utilizar datos para análisis secundario incluyen los siguientes:

- ¿Son válidos los datos disponibles?
- ¿Son confiables?
- ¿Son precisos?
- ¿Cuáles son las tasas de datos de respuesta y datos faltantes?

McCaston (2005) sugiere también revisar:

- Las credenciales de la organización que mantiene la base de datos.
- Los métodos y reglas de decisión utilizados para limpiar los datos.
- La antigüedad de la información en la base de datos (actual o desactualizada).
- Si los datos tienen sentido y son consistentes con los de otras fuentes.

Por ejemplo, considérese una unidad de evaluación que está pensando en utilizar datos de un gran sistema de información para la administración computarizado, con datos sobre créditos universitarios obtenidos por medio de un importante programa del gobierno. Como requerimiento de este programa, el ingreso de los estudiantes debe haber sido inferior a algún nivel requerido. Los evaluadores que consideran utilizar estos datos deben verificar que muestran que los estudiantes no elegibles para la financiación no la recibieron. Si encuentran que los estudiantes no elegibles recibieron financiación, deben determinar si el problema es de entrada errónea de datos (precisión) o implementación impropia de los requerimientos. Los evaluadores deben verificar también que a los estudiantes elegibles no se les negó la financiación, o que los estudiantes no recibieron más de un crédito. Los evaluadores pueden efectuar en primera instancia corridas de datos para determinar la naturaleza y grado de estos problemas en la base de datos. Para verificar éstos, pueden telefonear a una muestra de estudiantes con el fin de confirmar tanto que recibieron la financiación como que asistieron al colegio. Los evaluadores podrían además efectuar verificaciones con los colegios para determinar si los registros muestran que los estudiantes asistieron y recibieron financiación del programa.

Al utilizar datos secundarios deben seguirse los procesos siguientes:

- Averiguar qué se requiere para transferir los datos al computador propio. Algunas veces, la organización que mantiene la base de datos prefiere efectuar el análisis en lugar del evaluador y sólo existe esta opción. (Si es posible y práctico, es preferible a transferir el archivo.)
- Asegurar que no haya virus antes de transferir los datos al computador.
- Obtener la estructura de la base de datos, el diccionario de datos y los esquemas de codificación.
- Verificar la exactitud de los datos mediante pruebas y validaciones cruzadas.

001El camino.indd 290 08/11/2010 03:27:52 p.m.

- Transferir los datos en forma que se eviten nuevos errores de introducción (no ingresarlos nuevamente). Verificar manualmente algunos datos con el fin de asegurar que fueron transferidos como se planeó.
- Referenciar apropiadamente todos los datos secundarios (fecha de publicación, citas de autores apropiadas).

En el cuadro 8.4 se resumen las ventajas y problemas de utilizar datos disponibles.

Herramienta 3: observación

La observación puede constituir una poderosa herramienta de recolección de datos. Con sólo utilizar los ojos pueden observarse, por ejemplo, patrones de tráfico, patrones de uso de la tierra, diseño de entornos urbanos y rurales, calidad de viviendas, condición de vías y quién asiste a una clínica rural para servicios médicos.

La observación es una herramienta útil de recolección de datos cuando el propósito de la evaluación es recopilar datos para normas de referencia y descriptivos, o para documentar actividades, procesos y productos del programa. Es apropiada en una variedad de situaciones (cuadro 8.5).

Las observaciones pueden ser estructuradas o semiestructuradas:

- Las observaciones estructuradas determinan antes de la intervención precisamente qué se observará durante cuánto tiempo. Los observadores usualmente utilizan una lista de verificación para contar eventos o casos, o un cronómetro para actividades de tiempo.
- Las observaciones no estructuradas seleccionan el método, dependiendo de la situación, sin ideas preconcebidas o un plan sobre qué observar o durante cuánto tiempo. El observador mira y registra a medida que se desenvuelve la situación.
- Las observaciones semiestructuradas se efectúan cuando el evaluador tiene una idea general de qué observa, pero sin un plan específico. El evaluador puede simplemente anotar lo que encuentra interesante, típico, inusual o importante. El evaluador puede ocuparse en la anotación continua sobre transacciones cuando ocurren, o centrarse en acciones específicas.

Cuadro 8.4 *Ventajas y problemas de utilizar datos disponibles*

Ventajas	Los datos son usualmente menos costosos y se obtienen más rápido que al recopilar datos originales.
Problemas	Los datos pueden no ser exactamente lo que se requiere.
	Puede ser difícil mantener el acceso a los datos a largo plazo.
	Deben verificarse la validez y la confiabilidad de los datos, y corregirse los errores de codificación.

Fuente: los autores.

Observación estructurada

Utilizando un enfoque de observación estructurada, Stake (1995) desarrolló una forma de observación basada en temas (gráfico 8.2). En muchas formas de observación estructurada se utilizan abreviaturas y símbolos para permitir un mayor espacio en el formulario de recolección de datos. Estos formularios pueden modificarse para suplir las necesidades de evaluaciones particulares.

El gráfico 8.3 presenta otro ejemplo de un formulario utilizado para observación estructurada. Esta herramienta se utiliza para recopilar datos de observación sobre destrezas infantiles de comunicación.

Cuadro 8.5 *Razones para utilizar y ejemplos de observación*

Razón para utilizar la observación	*Ejemplos*
Obtener información directa.	Hacer visitas aleatorias a colegios, hogares, parcelas u otros sitios y observar en lugar de sólo preguntar.
	Observar operaciones y procedimientos en oficinas, colegios, hospitales u otros sitios en lugar de sólo depender de informes.
	Registrar discretamente número de ocurrencias de grupos étnicos, géneros o de edad en una actividad particular.
Comprender un comportamiento, proceso, situación en desarrollo o evento en curso.	Observar y descubrir lo que se hace en cada fase del proyecto o programa.
	Observar la interacción de los niños con otros niños en el patio de recreo, el trato de los padres con los hijos, de los maestros con los estudiantes, de los administradores con los empleados.
	Observar a los administradores dirigir reuniones de negocios antes y después de programas de capacitación de funcionarios.
Examinar evidencia física, productos o salidas.	Observar alimentos y otros artículos de venta en el mercado.
	Observar periódicamente la orilla de un lago objeto de un programa de limpieza.
	Hacer que un equipo de expertos inspeccione la calidad de verduras y legumbres en una hortaliza.
	Inspeccionar jardines, cartas noticiosas, libros de proyectos, etcétera.
Proveer una alternativa cuando los procedimientos de recolección por escrito u otros son impracticables o inapropiados.	Hacer que varios participantes se presten como voluntarios para observar e informar sobre la ejecución de un programa en lugar de hacer que los participantes llenen formularios.
	Observar la dinámica y las preocupaciones expresadas en un taller para nuevos inmigrantes.
	Hacer que los instructores observen las clases de los demás; anotar la dinámica, preguntas y nivel de participación.

Fuente: adaptado de University of Wisconsin Cooperative Extension, 1996.

Gráfico 8.2 *Muestra de formulario de observación basado en temas*

Observador:	Colegio:	Fecha:	Tiempo de observación: De Para
Maestro: Hombre Mujer	Edad 25 35 50 65	Grado:	Tiempo de escritura:
Experiencia del maestro: 0 - - Meses	Instrucción directa Alta - - Baja	N° Estudiantes:	Materia:
Sinópsis de la lección, actividades:		Comentarios sobre temas de educación en ciencias: 1 respuesta a cortes presupuestarios 2 lugar de la autoridad 3 preparación del maestro 4 materiales de práctica	

Descripción del salón	Orientación pedagógica	Finalidad del maestro	Referencia a
Lugar de aprendizaje B - - A	Texto B - - A	Didáctica A - - B	Método cient. 0 - - M
Lugar de ciencia B - - A	Prueba estand. B - - A	Heurístico B - - A	Tecnología 0 - - M
Lugar de compet B - - A	Resol de probl B - - A	philetic B - - A	Ética, relig. 0 - - M

Fuente: adaptado de Stake, 1995.

Observación semiestructurada

Una observación semiestructurada puede no tener un plan específico. En cambio, puede identificar qué clase de cosas observar (cuadro 8.6).

Registro de las observaciones

Las observaciones pueden registrarse por lo menos de tres maneras:

- *Guía de observación:* forma impresa que provee espacio para registrar observaciones. (Ejemplos de guías de observación se encuentran en Stake [1995], University of Wisconsin Cooperative Extensión [1996] y Yin [2003]).
- *Hoja de registro o lista de verificación*: forma utilizada para registrar observaciones en forma de sí o no o de escala de clasificación. Las hojas de registro se utilizan cuando se trata de observar artículos, actores o atributos específicos y observables.
- *Notas de campo:* Es la forma menos estructurada de registrar observaciones. Las observaciones se registran en un estilo narrativo, descriptivo, cuando el observador nota u oye algo importante.

Notas de campo: Descripciones de lo que se experimenta y observa, citas de las personas observadas, sentimientos y reacciones del observador a lo observado y percepciones e interpretaciones generadas en el campo.

Las *notas de campo* contienen "los datos actuales que se recopilan. Consisten en descripciones de lo que se experimenta y observa, citas de las personas observadas,

Gráfico 8.3 *Muestra de formulario de observación estructurada*

Indicador temprano de comunicación (ITC)

Early Childhood Research Institute on Measuring Growth and Development

Nombre o No. del niño _____ Fecha de la prueba_____(MM/DD/AA)
Duración de la prueba_____ - _____
Min Seg
Forma: Cambio en la condición de la vivienda o establo (ver lista abajo) _____
Codificador primario:_____ Asesor: _____
Ubicación (encerrar una en un círculo): Hogar Centro Otra
(explicar en las notas)
Lenguaje de la administración:_____
Si confiabilidad, nombre del codificador de confiabilidad_____
Notas:

	Gestos	Vocaliza-ciones	Palabras solas	Múltiples palabras
Empieza 0:00 Sec.	G	V	S	M
1:00 Seg.	G	V	S	M
2:00 Seg.	G	V	S	M
3:00 Seg.	G	V	S	M
4:00 Seg.	G	V	S	M
5:00 Seg.	G	V	S	M
6 min. Fin Total	G	V	S	M

Lista de condición
Intérprete
Herramienta de intervención de lenguaje
Intervención médica (por ejemplo, tubos)
Consultor de salud mental
Enseñanza del medio o eventual
Ninguno
Nutricionista
Otro
Terapeuta físico
Proveedor de servicios primarios
Enfermera registrada
Interacción sensible
Trabajadora social
Terapeuta de habla/lenguaje

(Continúa en la página siguiente)

(Continuación gráfico 8.3)

	Gestos	Vocaliza-ciones	Palabras solas	Varias palabras	Total
Codificador primario					
Codificador de confiabilidad					
Acuerdo					
Desacuerdo					
Porcentaje					

Porcentaje general

Acuerdo

A/A#D

Determinar confiabilidad:

1. Registrar puntajes de codificador primario en primera línea.
2. Registrar puntajes de codificador de confiabilidad en segunda línea.
3. Registrar número sobre el cual estuvieron de acuerdo en tercera línea.
4. Registrar número sobre el cual estuvieron en desacuerdo en cuarta línea.
5. Calcular porcentaje de acuerdo para cada categoría de elemento básico.
6. Calcular porcentaje general de acuerdo, utilizando puntajes totales.
7. Calcular porcentaje promedio de acuerdo en las categorías (sumar acuerdos y desacuerdos de las categorías: tercera y cuarta líneas).

Fuente: Early Childhood Research Institute on Measuring Growth and Development 2008.

Cuadro 8.6 *Usos de la observación semiestructurada*

Observación de	*Ejemplos*
Características de participantes (individualmente o como grupo)	Género, edad, profesión/vocación, vestuario, apariencia, grupo étnico
	Actitud hacia el sujeto, hacia otros, sobre sí mismo
	Destrezas, hábitos y conocimiento
Interacciones	Nivel de participación, interés
	Relaciones de poder, toma de decisiones
	Clima general para aprendizaje, resolución de problemas
	Niveles de apoyo, cooperación frente a conflicto
Comportamiento no verbal	Expresiones faciales
	Gestos
	Posturas
Líderes de programa, presentadores	Claridad de comunicación
	Destrezas de liderazgo de grupo, estimular participación
	Conciencia de clima de grupo
	Flexibilidad, adaptabilidad
	Conocimiento de un sujeto, uso de ayudas, otras técnicas de enseñanza o aprendizaje
	Secuencia de actividades
Alrededores físicos	Habitación (espacio, comodidad, adecuación)
	Cortesías (bebidas, etc.)
	Arreglos de asientos
Productos de un programa	Demostraciones, facilidad, plan
	Folletos, manuales, cartas noticiosas

Fuente: Cloutier y otros, 1987.

sentimientos y reacciones del observador a lo observado y percepciones e interpretaciones generadas en el campo" (Patton, 2002, p. 305).

Según Lofland (1971, p. 102), las notas de campo son "el determinante más importante de lograr un análisis cualitativo. Las notas de campo proporcionan la razón de ser del observador. Si… no se hacen [el observador] bien podría no estar en el entorno".

Patton (2002) discute técnicas de tomar notas de campo para las observaciones, y describe la importancia de describir todo lo que el observador cree que valga la pena anotar y no confiar nada al recuerdo. Tan pronto como le sea posible, el observador debe registrar cualquier información que haya contribuido a comprender el contexto, el entorno y lo que haya sucedido. Las notas de campo deben contener información descriptiva que durante el análisis le permita al evaluador regresar mentalmente a la

observación para experimentar esa observación. Patton sugiere registrar información básica; por ejemplo, dónde tuvo lugar la observación, quién estaba presente, cómo era el entorno físico, qué interacciones sociales ocurrieron y qué actividades se desarrollaron.

Anota la importancia de utilizar términos específicos en lugar de generales: palabras como *pobre, airado* e *inquieto* no son suficientemente descriptivas. Según Patton, "Tales palabras interpretativas ocultan lo que realmente sucedió y no revelan los detalles de la situación".

Capacitación y preparación de observadores

Patton (2002, p. 260) señala que solamente porque una persona esté "equipada con sentidos que funcionan no significa que esa persona sea un observador calificado", y discute la importancia de capacitar y preparar observadores, identificando seis componentes de la capacitación:

- Aprender a prestar atención, ver lo que hay que ver y oír lo que hay que oír.
- Practicar la escritura descriptiva.
- Adquirir disciplina en el registro de notas de campo.
- Saber separar el detalle de las trivialidades para lograr los primeros sin quedar abrumados por las segundas.
- Utilizar un método riguroso para validar y triangular las observaciones.
- Informar las fortalezas y limitaciones de la propia perspectiva, que requiere tanto autoconocimiento como autodivulgación.

En cuanto sea posible, debe utilizarse más de un observador. Todos los observadores deben capacitarse de tal modo que observen según los procedimientos acordados.

El procedimiento de recolección de datos de observación debe probarse con un programa piloto antes de utilizarse para la evaluación. Para hacer esto, por lo menos dos observadores deben ir a la misma zona y completar sus notas de campo con formas de codificación y hojas de clasificación. Después de completar las hojas, éstas deben compararse y si existen grandes diferencias, debe proveerse mayor capacitación y clarificación. Si las diferencias son pocas, puede adoptarse el procedimiento.

En el cuadro 8.7 se resumen las ventajas y problemas de recopilar datos de observaciones.

Cuadro 8.7 *Ventajas y problemas de la recolección de datos de observación*

Ventajas	La recolección es de datos de comportamiento real y no autoinformes o percepciones.
	La recolección es de datos en tiempo real y no retrospectivamente.
Problemas	El muestreo, el registro y la interpretación de los datos pueden ser difíciles.
	La recolección de datos puede ser intensiva en mano de obra.

Fuente: los autores.

Herramienta 4: encuestas y entrevistas

Las encuestas constituyen herramientas excelentes para recopilar datos sobre las percepciones, opiniones e ideas de las personas. Son menos útiles para medir el comportamiento, porque lo que dicen las personas puede no reflejar lo que realmente hacen.

Las encuestas pueden ser estructuradas o semiestructuradas, realizarse personalmente o por teléfono, o autorrealizarse, haciendo que las personas respondan en un formulario enviado por correo o por Internet. Las encuestas pueden hacerse a una muestra o a toda una población (para una discusión sobre el muestreo, *ver* el capítulo 9).

Encuestas estructuradas y semiestructuradas

Las *encuestas estructuradas* son aquellas en las que se incluye un rango de elecciones de respuesta, del cual el encuestado elige una o más. A todos los encuestados se les hacen en exactamente la misma forma las mismas preguntas, y para responderlas se les ofrecen exactamente las mismas elecciones.

¿Cuántas opciones de respuestas deben utilizarse? Para respuestas nominales, por ejemplo una región de un país, ocupación primaria, o grupo de edad, el número de respuestas debe cubrir todas las respuestas posibles. Cuando se utilicen escalas para indicar respuestas, el número de opciones de respuestas debe ser en general un número impar (3, 5 o 7), de manera que la respuesta neutral sea evidente para el encuestado. (Algunas veces se utilizan escalas de números pares, para que el encuestado elija entre una clasificación "satisfactoria" o "insatisfactoria"). En las encuestas semiestructuradas se presenta el mismo conjunto general de respuestas, pero se permiten respuestas abiertas a todas o a la mayoría de las preguntas (recuadro 8.6).

Encuesta estructurada: Encuesta que incluye un rango de elecciones de respuesta, del cual el encuestado elige una o más.

Las encuestas semiestructuradas son las que presentan preguntas predominantemente abiertas y son especialmente útiles cuando el evaluador desea obtener un conocimiento más profundo de las reacciones a experiencias, o comprender las razones de por qué los encuestados mantienen actitudes particulares. Las encuestas semiestructuradas deben tener un propósito definido claramente. Con mucha frecuencia, es más práctico entrevistar personas sobre los pasos de un proceso, los papeles y responsabilidades de varios miembros de una comunidad o equipo, o una descripción de cómo funciona un programa, que intentar desarrollar una encuesta escrita que capture todas las variaciones posibles. Si los encuestados potenciales sienten armonía y se les plantean preguntas interesantes, pueden estar dispuestos a ser entrevistados por una hora o más. En contraste, pocas personas estarían dispuestas a emplear una hora para llenar un cuestionario. Idealmente, dos personas deberían realizar cada entrevista, para que puedan comparar notas. El empleo de dos entrevistadores puede ayudar también a resolver disputas sobre lo que se dijo.

Encuestas semiestructuradas: Son las que presentan preguntas predominantemente abiertas.

Recuadro 8.6 *Preguntas de encuestas estructuradas y semiestructuradas*

Ejemplos de preguntas estructuradas son las siguientes:

1. ¿Ha sido útil este taller para ayudarle a aprender a evaluar su programa?
   - Poco o nada.
   - En alguna medida.
   - En medida moderada.
   - En gran medida.
   - En muy gran medida.
   - Sin opinión.
   - No se aplica.

2. ¿Existe una fuente de agua limpia para todas las personas de la aldea a menos de 500 metros de su hogar?
   - Sí.
   - No.

Ejemplos de preguntas semistructuradas son:

1. ¿Qué aprendió del taller de evaluación de programas que haya utilizado en su trabajo?
2. ¿Cuáles son las fuentes de agua limpia para esta aldea?

En el cuadro 8.8 se destacan las ventajas de las encuestas estructuradas y semiestruturadas.

Cuadro 8.8 *Ventajas de las encuestas estructuradas y semiestructuradas*

Encuestas estructuradas	*Encuestas semiestructuradas*
Difíciles de desarrollar: es necesario que la encuesta cubra todas las piezas posibles de información.	Más fáciles de desarrollar: la encuesta puede incluir preguntas amplias y abiertas que capturan lo que haya faltado en las secciones estructuradas, reduciendo el peligro de dejar algo por fuera.
Más fáciles de completar: marcar un cuadro toma menos tiempo que escribir una respuesta narrativa.	Más difíciles de completar: fatigoso para las personas completar cuestionarios autorrealizados.
Más fáciles de analizar.	Difíciles de analizar, pero suministran una fuente de datos más rica; interpretación de respuestas abiertas sujeta a sesgo.
Más eficientes cuando se trabaja con grandes números de personas.	

Fuente: los autores.

Modos de realización de encuestas

Los medios más frecuentes para realizar encuestas son:

- Cuestionarios telefónicos.
- Cuestionarios distribuidos por correo, correo electrónico o Internet.
- Cuestionarios administrados (comunes en el contexto del desarrollo).

Las encuestas telefónicas son útiles para conocer experiencias, opiniones o descripciones personales individuales de un proceso. Las encuestas personales y telefónicas son similares y solamente difieren en el medio en donde se realizan.

Los cuestionarios pueden entregarse personalmente a los encuestados o por correo, correo electrónico o Internet. Los encuestados responden las preguntas en un formulario y luego retornan el formulario lleno al entrevistador. Las encuestas autorrealizadas deben ser cortas y no tomar más de 20 minutos (mientras menos tiempo, mejor). De todos los enfoques, la investigación sugiere que es más probable que las personas den respuestas honestas a preguntas sensibles en una encuesta anónima autorrealizada.

El uso del correo, el correo electrónico o la Internet es mucho menos costoso que realizar encuestas personales. Las encuestas pueden enviarse a cualquier zona en la que los encuestados potenciales tengan acceso a estos métodos de distribución. Si en una zona no se reciben envíos postales regularmente o se tiene capacidad limitada de telefonía o Internet, no se tendría una buena participación en la encuesta y se produciría una fuente importante de sesgo. Los bajos niveles de alfabetización hacen que estos enfoques sean inapropiados en muchos contextos de desarrollo.

Las encuestas administradas, usuales en el contexto del desarrollo, pueden ser estructuradas, semiestructuradas o mixtas. Estas encuestas escritas son leídas por un recopilador de datos, quien registra las respuestas. Por lo general, es preferible utilizar preguntas cerradas, pero incluir una o dos abiertas al final. Estas preguntas abiertas a menudo hacen que las personas se sientan cómodas, pues pueden agregar cualquier cosa que consideren que omitió la encuesta o comentar sobre la misma. Una o dos preguntas abiertas son todo lo que se requiere y más puede ser fatigoso. Puede requerirse mucho tiempo para analizar las respuestas a estas preguntas abiertas, pero proporcionan percepciones y citas útiles para ilustrar temas importantes de la evaluación.

En las *Encuestas telefónicas asistidas por computador* (ETAL) se utiliza un sistema de computador interactivo que le ayuda al entrevistador a hacer las preguntas por teléfono. El entrevistador inicia la entrevista y el programa de computador controla las desviaciones o los saltos de preguntas. La elección de las preguntas se basa en las respuestas a otras preguntas, permitiendo entrevistas más personalizadas y sofisticadas de lo que es posible con cuestionarios en papel. Durante la entrevista, los entrevistadores ingresan los datos, junto con una codificación simple al sistema del computador. La mayoría de las preguntas se hace en formato de elección múltiple, de modo que el entrevistador simplemente hace clic en la respuesta

ETAL: Entrevista en la que se utiliza un sistema de computador interactivo para ayudarle al entrevistador a hacer las preguntas por teléfono.

correcta; el sistema de computador convierte los datos a un código y los almacena en una base de datos (Unescap, 1999).

En el cuadro 8.9 se resumen las ventajas y los problemas de utilizar encuestas.

Desarrollo de las preguntas de la encuesta

El desarrollo de las preguntas de la encuesta implica:

- Seleccionar las preguntas.
- Redactar las preguntas y respuestas.
- Elaborar la secuencia de las preguntas.
- Elegir un diseño para la encuesta.
- Revisar, traducir y probar preliminarmente las preguntas.

Seleccionar las preguntas. Los evaluadores pueden utilizar preguntas abiertas o cerradas. Las preguntas abiertas no pueden responderse con una respuesta o una elección simple. Con mucha frecuencia los encuestados se sienten incómodos o indispuestos a responder por escrito preguntas abiertas.

Las preguntas cerradas pueden responderse con una simple pieza de información. La pregunta "¿Cuál es la fecha de su nacimiento?" es una pregunta cerrada porque puede responderse sólo con la fecha de nacimiento, sin detalles; solo una simple pieza de información. Una segunda forma de pregunta cerrada es una pregunta dicotómica, como las preguntas con respuestas de sí o no, o verdadero o falso. Una tercera forma de pregunta cerrada es una pregunta de selección múltiple. Muchos expertos aconsejan utilizar principalmente preguntas cerradas e incluir una o dos abiertas al final de la encuesta (Burgess, 2000) o dejar un espacio para un comentario después de cada pregunta.

Utilizar muchas formas distintas de preguntas puede causar confusión al encuestado. Jackson y Agrawal (2007) sugieren utilizar no más de tres formas de preguntas en la mayoría de las encuestas.

De acuerdo con el Living Standards Measurement Survey Study (LSMS) (Banco Mundial, 1996), es útil desarrollar preguntas de encuestas por niveles (cuadro 8.10). En esta forma se obtiene un cuestionario bien balanceado.

Cuadro 8.9 *Ventajas y problemas de las encuestas*

Ventajas	Los encuestados indican lo que piensan, creen o perciben.
Problemas	Las personas pueden no recordar exactamente su comportamiento, o pueden estar reacias a revelar su comportamiento, particularmente si es ilegal o estigmatizado.
	Lo que las personas piensan que hacen o dicen no es siempre lo mismo que lo que realmente hacen.

Fuente: los autores.

Cuadro 8.10 *Niveles de refinamiento para determinar el contenido del cuestionario*

Nivel	*Descripción*
Definir los objetivos generales	Definir objetivos (estudiar la pobreza, comprender los efectos de las políticas oficiales en las familias).
Determinar el equilibrio entre las secciones	Definir los temas más importantes (incidencia de subsidios de precios a los alimentos, efectos de cambios en la accesibilidad o costo de los servicios de salud y educación del gobierno, efecto de cambios en el clima económico como resultado de ajuste estructural o transición de una economía de planificación central a una economía de mercado.
Determinar el equilibrio al interior de las secciones	Por ejemplo, dentro del sector educativo, definir cuáles de los siguientes son más importantes para el país en el momento: los niveles y determinantes de las inscripciones, la escasa asistencia, el aprendizaje y las diferencias en los indicadores masculinos y femeninos, el impacto del número de años de escolaridad en los ingresos en el sector formal y en la agricultura, y cómo difieren o si difieren; cuáles niños tienen textos o reciben almuerzos escolares; cuánto deben pagar los padres por la escolaridad.
Redactar preguntas para estudiar asuntos o programas específicos	Por ejemplo, en un estudio sobre acceso a textos, el evaluador necesita saber cuántas materias cubren los textos; si serán para estudiantes individuales o compartidos; si se podrán llevar a casa o sólo se utilizarán en el salón de clases; si se utilizarán sólo para un año o para varios; si deberán pagarlos los estudiantes o se suministrarán sin cargo; el momento en que se espera estarán disponibles; si los adquiridos en librerías son mejores o peores que los suministrados por el colegio.

Fuente: Banco Mundial, 1996.

Distinguir el tipo de pregunta obliga al evaluador a ser claro sobre lo que se pregunta, y le ayuda al encuestado a responder apropiadamente. Patton (2002) describe seis clases de preguntas:

- *Preguntas de experiencia y comportamiento*: qué hace o espera hacer una persona, para producir como respuesta comportamientos, experiencia, acción y actividades.
- *Preguntas de opinión y valores*: dirigidas a comprender los procesos cognitivos e interpretativos de las personas (no sus acciones y comportamientos).
- *Preguntas de sentimientos*: dirigidas a despertar las emociones de las personas a experiencias y pensamientos, buscando respuestas adjetivas, como ansioso, feliz, temeroso, intimidado y confiado. La pregunta debería indicarle a la persona entrevistada que se le solicitan opiniones, creencias y juicios considerados y no sentimientos.
- *Preguntas de conocimiento*: preguntas para indagar sobre la información factual del encuestado: qué sabe.
- *Preguntas sensoriales*: preguntas sobre lo que se ve, oye, toca, gusta y huele.
- *Preguntas de antecedentes o demográficas*: edad, educación, ocupación y otras características de la persona entrevistada.

001El camino.indd 302 08/11/2010 03:27:53 p.m.

Redactar las preguntas y respuestas. Cada uno de los seis tipos de preguntas de Patton puede hacerse en tiempo presente, pasado o futuro. Por ejemplo, un evaluador puede preguntar:

- ¿Qué sabía usted sobre el tratamiento para VIH/Sida hace cinco años? (tiempo pasado).
- ¿Qué aprendió hoy de la presentación sobre VIH/Sida? (tiempo presente).
- ¿Qué le gustaría aprender sobre VIH/Sida? (tiempo futuro) (Patton, 2002).

Las preguntas de evaluación deben ser claras (recuadro 8.7), pues, de no ser así, el encuestado puede sentirse incómodo, ignorante, confundido u hostil. Los evaluadores deben aprender los términos utilizados por los encuestados cuando se refieren al proyecto, el programa o la política, y deben utilizar estas palabras al elaborar la encuesta.

El evaluador debe tener cuidado de no llevar al encuestado a dar una respuesta deseada. Por ejemplo, considérense las preguntas siguientes:

> "¿En qué grado se han incrementado sus beneficios, de haber sucedido esto, desde la capacitación?"
> "Si se han incrementado sus beneficios en un grado ligero, moderado, sustancial o grande, ¿lo atribuye a la capacitación?"

Claramente, estas preguntas "llevan al testigo" hacia donde quiere llegar el evaluador. Sería mejor preguntar sobre las tendencias de los beneficios en un período mayor de tiempo y luego preguntar, si hubo un incremento, a qué se atribuyó.

Recuadro 8.7 *Consejos para redactar preguntas de encuestas*

Los consejos siguientes pueden ayudar a los evaluadores a redactar preguntas, de modo que los encuestados conozcan la información que se les solicita:

- Utilizar términos sencillos que tengan el mismo significado para todos los encuestados.
- Elaborar categorías específicas de preguntas y respuestas.
- Evitar opciones de preguntas y respuestas de doble sentido. (Evitar preguntas y respuestas en que se utilicen las palabras "y" o "o").
- Evitar preguntas que supongan conocimiento. De ser necesario, suministrar la información pertinente en la introducción a la pregunta.
- Estar prevenido sobre los dobles negativos en las combinaciones pregunta-respuesta (por ejemplo, evitar el uso de "no" en preguntas con respuesta de sí o no).
- Hacer que las opciones de respuestas sean mutuamente excluyentes.
- Hacer que las opciones de respuestas sean equilibradas.
- Evitar preguntas sujetas a objeciones, impertinentes o transigentes.

Fuente: TC Evaluation Center, 2007. Para otras fuentes útiles, *ver* Foddy (1993), Fowler (2002) y Sue y Ritter (2007).

Los cambios ligeros en la forma de redactar las preguntas pueden tener un impacto significativo en la forma como responda la gente.

> Varios investigadores han observado los efectos de modificar adjetivos y adverbios. Palabras como *usualmente, a menudo, algunas veces, ocasionalmente, pocas veces y rara vez*, se utilizan 'comúnmente' en los cuestionarios, aunque es claro que no significan lo mismo para todas las personas. Algunos adjetivos tienen una alta variabilidad y otros, baja. Las siguientes [expresiones] tienen significado altamente variable y deben evitarse en las encuestas: *un mandato claro, muy, numeroso, una mayoría sustancial, una minoría de, muchos, un número considerable de* y *varios*. Otras [expresiones] producen menor variabilidad y generalmente tienen un significado más compartido. Estas son: *mucho, casi todos, virtualmente todos, la mayoría de, un consenso de, un pequeño número de, no demasiados, casi ninguno, difícilmente alguno, un par* y *unos pocos* (StatPac, 2007).

Frary (1996) ofrece varias sugerencias para diseñar cuestionarios eficaces. Entre sus sugerencias para respuestas clasificadas están las siguientes:

- Ordenar las respuestas en orden ascendente de izquierda a derecha: *1) Nunca, 2) Pocas veces, 3) Ocasionalmente, 4) Con frecuencia*.
- Considerar combinar categorías de respuestas si es muy improbable que los encuestados marquen "nunca" y si "pocas veces" tuviese la connotación de un nivel casi equivalente de actividad: *1) pocas veces o nunca, 2) ocasionalmente, 3) con frecuencia*.
- Solicitar a los encuestados clasificar estímulos tanto positivos como negativos (respuestas de ambos extremos de la escala). En esta forma, los encuestados deben evaluar cada respuesta y no estar de acuerdo o en desacuerdo uniformemente con todas las respuestas.
- De ser factible, reducir el número de respuestas posibles. Por ejemplo, si es probable que las opiniones de los encuestados sean claras, utilizar un simple *1) de acuerdo, 2) en desacuerdo*. Cuando muchos encuestados tienen opiniones no fuertes o bien formadas, utilizar más opciones: *1) de acuerdo, 2) tiende a estar de acuerdo, 3) tiende a estar en desacuerdo, 4) en desacuerdo.*

Después de redactar las preguntas, el evaluador debería leer cada pregunta y respuesta para verificar el leguaje y la lógica. ¿Es correcta la gramática? ¿Sigue la respuesta a la pregunta en forma lógica?

Es difícil redactar preguntas para encuestas porque deben ser comprensibles para todos, y las palabras tienen varios significados y connotaciones. Si se hace una pregunta en forma que las personas no la entienden o la entienden de varias formas, estarán respondiendo una pregunta distinta. Existe también el riesgo de recopilar datos inútiles. Por ejemplo, un jefe de una agencia desea averiguar qué tanta capacitación ha recibido el personal de su organización en computadores, y podrían hacerse estas preguntas:

001El camino.indd 304 08/11/2010 03:27:53 p.m.

- ¿Ha recibido alguna capacitación en los últimos tres meses?
- ¿Ha recibido alguna capacitación en los últimos seis meses?
- ¿Ha recibido alguna capacitación en el último año?

Con este conjunto de preguntas el problema es que quien recibió capacitación en los últimos tres meses, responderá "sí" a las tres preguntas. Una pregunta mejor diseñada sería "¿Cuántos cursos de capacitación ha recibido en cada uno de los períodos siguientes?

a. En los últimos 1 a 3 meses: _____
b. En los últimos 4 a 6 meses: _____
c. En los últimos 7 a 9 meses: _____
d. En los últimos 10 a 12 meses: _____
e. En los últimos 13 a 24 meses: _____

Las preguntas redactadas con deficiencia pueden frustrar a los encuestados, haciendo que conjeturen las respuestas o incluso arrojen la encuesta a la basura. Dado que las preguntas mal hechas no pueden guardarse en el análisis, deben estar bien redactadas. Por consiguiente, los evaluadores deben separar abundante tiempo para revisar y hacer pruebas preliminares de la encuesta.

Deben garantizarse y protegerse tanto la confidencialidad como el anonimato. La confidencialidad es un principio ético que protege la información, permitiendo su acceso sólo a personal autorizado. El anonimato significa que la identidad o la información personal sobre una persona no será compartida. Muchos evaluadores incluyen una declaración semejante a "Sus respuestas serán tratadas con confidencialidad y en todo momento los datos se presentarán de tal forma que no pueda conectarse su identidad con datos específicos publicados" (Burgess, 2001).

Elaborar la secuencia de las preguntas. Elaborar la secuencia de las preguntas en el orden siguiente puede mejorar las posibilidades de que se llene una encuesta:

- Empezar con preguntas de interés para el encuestado.
- Mantener el flujo lógico de las preguntas y evitar desviaciones complejas.
- Agrupar preguntas similares o relacionadas e intentar establecer una secuencia lógica para las preguntas (Burgess, 2001).
- Pasar de lo fácil e interesante a lo difícil e incómodo, de lo general a lo específico (Jackson y Agrawal, 2007).
- Escribir las preguntas en orden cronológico (Jackson y Agrawal, 2007).
- Hacer las preguntas más importantes hacia las dos terceras partes de la encuesta, pues algunos encuestados no podrán terminar la encuesta (Jackson y Agrawal, 2007).
- Situar las preguntas personales o confidenciales al final de la encuesta; si se sitúan al principio, algunos encuestados no continuarán (Frary, 1996).
- Fijar prioridades en las preguntas y considerar abandonar las de baja prioridad (Jackson y Agrawal, 2007).

001El camino.indd 305 08/11/2010 03:27:53 p.m.

Jackson y Agrawal (2007) sugieren que la mayoría de las encuestas escritas no sean de más de una página. Para encuestados con educación universitaria, pueden ser de dos.

Elegir un diseño para la encuesta. El diseño de la encuesta impresa merece alguna consideración:

- Debe tener título y fecha de versión.
- Debe ser despejada y sin encabezados innecesarios.
- Debe incluir información de contacto y devolución.
- Debe incluir una introducción breve (en caso de que la carta de cubierta esté desubicada) que indique el propósito de la encuesta y explique cómo debe utilizarse.
- Las elecciones de preguntas y respuestas deben ser formateadas y redactadas en forma consistente. Cada pregunta debe estar numerada.
- La encuesta debe incluir instrucciones sobre cómo deben responderse las preguntas, y las instrucciones deben indicar cuántas respuestas pueden elegirse (sólo una o las que se apliquen) y cómo indicar las selecciones (marque el recuadro, encierre en círculo, dé una respuesta corta) (Burgess, 2001).

Jackson y Agrawal (2007) sugieren colocar los recuadros para marcar, o líneas de respuesta con suficiente espacio entre ellas, de modo que el evaluador pueda identificar la selección hecha. También tratan la importancia de darle a la encuesta una imagen profesional.

Una carta de cubierta debe acompañar la encuesta. Esta carta identifica al promotor y suministra a los participantes información sobre el propósito del cuestionario. Jackson y Agrawal (2007) sugieren que en la carta se identifique lo siguiente:

- El propósito de la encuesta (y por qué le interesa llenarla al encuestado).
- Cómo fue seleccionada la persona para la encuesta.
- Cómo se utilizarán los datos.
- Si se mantendrá el anonimato.
- El tiempo requerido para llenar la encuesta.
- Incentivos, si los hay.
- Información de contacto para mayor información o clarificación.

Si la encuesta se envía por correo, debe incluirse un sobre estampillado con la dirección del encuestador, u otro procedimiento establecido para devolver la encuesta llena.

Revisar, traducir y probar preliminarmente las preguntas. El proceso de desarrollo de encuestas es iterativo. Una vez propuesta la primera versión de la encuesta, las partes interesadas deben revisar detenidamente la propuesta. Durante la revisión, deben tomar notas y compartir críticas y sugerencias. Luego se hacen correcciones a la encuesta. Este proceso debe repetirse varias veces antes de estar lista la encuesta para su prueba preliminar.

Cuando se requiere más de un idioma para recopilar los datos, debe traducirse la encuesta. Los responsables de ésta deben asegurarse de que los instrumentos hayan sido traducidos con exactitud. Al traducir encuestas, una persona capacitada, que conozca ambos idiomas y esté familiarizada con el tema del cuestionario, debería hacer la primera traducción (*ver* Behling y Law, 2000). Una vez traducida la encuesta al idioma local, otra persona debe traducirla de nuevo al original. Con este proceso se verifican vacíos o malentendidos en lo traducido. Esta es la única forma de asegurar una traducción exacta. La falta de acuerdo en las partes de la encuesta debe conciliarse antes de efectuar cualquier prueba piloto.

En la mayoría de los casos, las encuestas se imprimen sólo en el(los) idioma(s) oficial(es) del país, y se emplean equipos de entrevistadores con destrezas en comunicación en varios idiomas locales para recopilar los datos. En estos casos, unas pocas preguntas o frases básicas se traducen a los idiomas locales y se presentan en el manual de la encuesta. Pueden requerirse intérpretes locales para encuestar personas que hablen otros idiomas.

Las preguntas de encuestas siempre deben redactarse en términos sencillos en el idioma que se habla habitualmente. Las preguntas no deben escribirse en un estilo de lenguaje académico o formal. En idiomas locales, el vacío entre los idiomas hablado y escrito y la dificultad de equilibrar la sencillez con la precisión puede ser grande. Esto es válido especialmente para los idiomas que no se utilizan de manera habitual por escrito.

Una vez se ha llegado a un acuerdo sobre la encuesta, debe someterse a prueba de campo o piloto con un pequeño número de personas. Según los resultados de la prueba de campo, será probable que se requieran correcciones adicionales. Si la primera prueba sugiere muchos cambios, puede necesitarse otra prueba de campo que cubra parte o toda la encuesta, especialmente si hubiese problemas con varias traducciones.

Con todas las encuestas y entrevistas es esencial efectuar pruebas de campo o piloto (preprueba) con el instrumento de recolección de datos pronto. Esto significa solicitar a una muestra pequeña, pero representativa, de encuestados potenciales que llenen la encuesta y resalten las partes que necesiten clarificación. Una estrategia útil es sentarse con los encuestados mientras responden la encuesta, y pedirles que razonen en voz alta al llenarla. Este ejercicio puede aportar percepciones sobre la interpretación que los encuestados potenciales darán a las preguntas. A menudo, cosas consideradas tan claras como el cristal resultan ser confusas o ambiguas en formas imprevistas. Si los encuestados entienden mal una pregunta, el evaluador debe corregirla y volver a probarla antes de implementar el instrumento.

La meta de una prueba de campo es el de asegurar que la encuesta sea capaz de recopilar la información que se intenta. Una buena prueba piloto examina la encuesta en tres niveles:

- *Como un todo*: ¿Son consistentes todas las partes de la encuesta? ¿Existen partes en las que se hace la misma pregunta?

- *Cada sección*: si la encuesta tiene más de una sección, ¿se recoge la información deseada en la información que se pide en cada sección? ¿Se da cuenta de todas las actividades principales? ¿Existen preguntas que no son pertinentes?
- *Preguntas individuales*: ¿Es clara la redacción? ¿Permite respuestas ambiguas la pregunta? ¿Existen interpretaciones alternativas?

Al efectuar una prueba de campo o piloto, es importante poner a prueba la encuesta entre personas de diversas zonas y todos los grupos principales de idioma y socioeconómicos. Esto puede incluir residentes rurales y urbanos, empleados en los sectores formal e informal y personas que hablen todos los grupos básicos de idiomas.

Si se efectúa la prueba sólo en un idioma, usualmente se tomará alrededor de una semana completarla. Si se va a realizar la encuesta definitiva en más de un idioma, la prueba piloto tomará más tiempo, pues debe hacerse la prueba de campo con una versión de cada idioma.

Al final de la prueba de campo, deben separarse una o dos semanas para revisar los resultados. El(los) equipo(s) que trabaje(n) en las pruebas de campo debe(n) reunirse para discutir y acordar los cambios necesarios. En una encuesta a gran escala usualmente se efectúan tres ciclos de pruebas piloto.

En los recuadros 8.8 y 8.9 se presentan lineamientos para efectuar encuestas. En el primero se enfatiza en el proceso de administración, y en el segundo se hace enfoque en la escritura de buenas preguntas.

Tarjetas de informe de ciudadanos

Tarjeta de informe de ciudadano: Tipo de encuesta que mide el funcionamiento de los servicios públicos.

Algunas veces llamada *tarjeta de puntaje de ciudadano*, es un tipo de encuesta con que se mide el funcionamiento de los servicios públicos, y se está utilizando con mayor frecuencia en los programas para el desarrollo. Por ejemplo, en Bangalore, India, un grupo de ciudadanos privados la utiliza para tratar problemas con los servicios públicos (Paul y Thampi, 2008). Las tarjetas de informe de ciudadanos se han utilizado por muchos años en los Estados Unidos para medir el desempeño de las agencias públicas.

Las tarjetas de informe de ciudadanos se utilizan para:

- Recopilar retroalimentación ciudadana sobre los servicios públicos de los usuarios reales del servicio.
- Evaluar el desempeño de los proveedores individuales de los servicios, o comparar su desempeño.
- Generar una base de datos de retroalimentación sobre los servicios que luego pueda ponerse a disposición del público (BAsD y ADBI, 2004).

Recuadro 8.8 *Lineamientos generales para efectuar encuestas*

Los lineamientos generales para efectuar encuestas son los siguientes:

- Hacerla sencilla, clara, fácil y corta.
- Ubicar otras personas que hayan realizado la clase de evaluación en la que está interesado. Encontrar y estudiar encuestas similares a la que piense que desea realizar.
- Asegurarse de que la gente sepa por qué les solicita participar (pero de manera que sea improbable sesgar sus respuestas).
- Hacer preguntas fáciles de responder y no frustrar el deseo del encuestado de ser claro.
- No solicitar a los encuestados información que les exija ir a un archivo u otra fuente. Si tiene que hacer esto, debe avisarlo por adelantado, de manera que el material pueda estar listo antes de entregar la encuesta.
- Respetar la privacidad de los encuestados. Tratar las encuestas en forma confidencial y con procedimientos para garantizar la privacidad. Nunca prometer confidencialidad a menos que pueda cumplirse.
- Respetar el tiempo y la inteligencia de los encuestados.
- Informar a los encuestados cómo fueron seleccionados, y por qué es importante su participación.
- No perjudicar: mantener las respuestas en forma confidencial. En su informe, utilizar respuestas agregadas. Asignar un número de identificación a los datos y destruir el vínculo con el nombre de la persona.
- Permitir que los encuestados sepan si recibirán una copia del informe final que utiliza los resultados del cuestionario.

Fuente: los autores.

Recuadro 8.9 *Consejos para escribir cuestionarios*

Los consejos siguientes pueden ayudar a los evaluadores a escribir cuestionarios:

1. Si es posible, utilizar como guía un cuestionario existente y modificarlo según sea necesario para adecuarlo a la situación. Es más fácil modificar que crear uno a partir de cero.
2. Seguir los lineamientos básicos siguientes:

    - Utilizar un lenguaje sencillo y claro apropiado para los encuestados.
    - Hacer sólo una pregunta a la vez; por ejemplo: "¿Hasta qué punto es claro el material, si lo es?" en lugar de "¿Hasta qué punto es claro y útil el material, si lo es?" (Si el material es claro, pero no útil, no hay forma de que la persona pueda dar una respuesta precisa).
    - Escriba sus preguntas de modo que todos los encuestados consideren sus respuestas aceptables. Conduzca a su pregunta ofreciendo el rango: "¿Hasta qué punto, si", o "cuán importantes o no son…?"

(Continúa en la página siguiente)

(Continuación recuadro 8.9)

- Proporcione categorías de respuestas apropiadas que sean mutuamente excluyentes. Si se solicitan grupos de edades, hacer categorías: 21-30, 31-40, 41-50, en lugar de 20-30, 30-40, 40-50.
- En cuanto sea posible, escribir las preguntas de modo que las respuestas vayan de lo negativo a lo positivo:

"¿Hasta qué punto fue… inútil ←…→ muy útil?, si lo fue".

- Evitar respuestas de tipo sí o no cuando la respuesta no es dicotómica. Más bien, intentar capturar una variedad de opiniones, solicitando a los encuestados responder a lo largo de una escala. Por ejemplo, proporcionar una escala de cinco puntos que vaya desde "poco o ningún grado" a "muy alto grado".
- Evitar los absolutos en cada extremo de la escala (pocas personas son absolutas sobre algo). Suavizar los extremos, utilizando "siempre o casi siempre" en un extremo de la escala, y "nunca o casi nunca" en el otro.
- Hacer preguntas sobre la situación actual, pues la memoria mengua con el tiempo.
- Permitir categorías de salida (utilizar "sin base para juzgar" y "sin opinión"). Si no se suministra una salida, los encuestados pueden dar respuestas sin sentido.
- Evitar el uso de dobles negativos.

3. Hacer que la encuesta sea fácil de completar. Provea recuadros para marcar e instrucciones suficientes para que los encuestados sepan lo que van a hacer: indique "marque sólo uno" o "marque los que se apliquen", según sea apropiado.
4. Haga primero las preguntas generales, después las más específicas y luego dos preguntas abiertas, por ejemplo: "¿Algún comentario sobre algo más que deberíamos saber?"
5. Solicitar sólo la información demográfica que se piense utilizar. Tener en cuenta que algunas personas pueden ser identificadas por su demografía. Coloque las preguntas demográficas hacia el final de la encuesta.
6. Haga que los expertos revisen las preguntas propuestas.
7. Si deben traducirse las preguntas, hacerlo y luego volverlas a traducir al idioma original para verificar la traducción.
8. ¡Haga pruebas preliminares, pruebas preliminares, pruebas preliminares! Haga que encuestados usuales respondan el cuestionario en lugar de sólo leerlo. Luego regrese a cada pregunta para obtener retroalimentación de ellos. ¿Es clara cada pregunta? ¿Comprendieron lo que se preguntaba? ¿Hay palabras desconocidas o frases no claras? ¿Existe una forma mejor de hacer cada pregunta?

Fuente: los autores.

En el cuadro 8.11 se muestra un ejemplo de tarjeta de informe de ciudadanos que informa sobre la satisfacción general con los servicios.

En los informes ciudadanos se recopila retroalimentación sobre los factores siguientes:

- Disponibilidad del servicio.
- Acceso al servicio.
- Confiabilidad del servicio.
- Calidad general del servicio.
- Satisfacción con el servicio.
- Grado de reacción del proveedor del servicio.
- Costos ocultos.
- Corrupción y sistemas de apoyo.
- Disposición a pagar (Banco Mundial, 2006).

Las tarjetas de informe de ciudadanos pueden utilizarse en muchas formas (cuadro 8.12), y también presentarse en diferentes niveles del gobierno.

En el cuadro 8.13 se resumen las ventajas y problemas de las tarjetas de informe de ciudadanos para la recolección de datos.

Cuadro 8.11 Resultados de tarjetas de informe de ciudadanos sobre satisfacción con servicios de infraestructura

Agencia	*Número de usuarios entre encuestados (miles)*	*Porcentaje de usuarios satisfechos*	*Porcentaje de usuarios insatisfechos*
Energía	1.024	43	15
Agua	775	41	19
Teléfono	203	87	12
Policía	98	45	36

Fuente: los autores.

Cuadro 8.12 Usos de las tarjetas de informe de ciudadanos

Tipo de estudio o proyecto	*Ejemplo*
Estudios de satisfacción con servicios urbanos	Siete ciudades de la India; Kampala, Uganda
Estudios de satisfacción provincial y nacional con la prestación de los servicios	India, Filipinas, Uganda
Estudios sectoriales	Hospitales públicos en Bangalore, India
Evaluaciones de programas	Seguridad alimentaria rural en Tamil Nadu, agua y saneamiento rural en Maharashtra, India
Proyectos de reforma de gobernanza	Bangladesh, Perú, Filipinas, Sri Lanka, Ucrania, Vietnam

Fuente: Banco Mundial, 2006.

Cuadro 8.13 *Ventajas y problemas de las tarjetas de informe de ciudadanos para la recolección de datos*

Ventajas	• Suministran retroalimentación sobre el desempeño.
	• Estructuradas para comunicación sencilla.
	• Contribuyen a reducir el sesgo en la recolección de datos.
Problemas	• Requieren capacidad para comprender esta clase de evaluación.
	• Se limitan a comparaciones entre servicios.
	• Requieren muestras grandes para poblaciones heterogéneas para servicios utilizados sólo por un pequeño porcentaje de la población.
	• Carecen de posibilidad de predicción sobre cómo responderán los funcionarios del gobierno.

Fuente: los autores.

Entrevistas

Una entrevista es un diálogo entre un entrevistador diestro y una persona entrevistada. La meta es producir información que pueda utilizarse para responder preguntas de evaluación (recuadro 8.10). La calidad de la información obtenida depende en gran parte de la destreza y personalidad del entrevistador (Lofland y Lofland, 1995). La clave de una buena entrevista es ser un buen escucha y un buen interrogador.

Cada cultura tiene sus propios valores y costumbres: una pregunta o gesto puede ofender a personas de una cultura y no de otra. Por consiguiente, la competencia cultural es una necesidad para el entrevistador, quien debe prestar atención a las reacciones de los encuestados durante la entrevista a fin de asegurar que no cause ofensa inadvertidamente (recuadro 8.11).

Al tratar con personas de todas las culturas, los entrevistadores deben tener presente lo siguiente:

- Toda persona aprecia ser tratada con respeto.
- Aun las personas que provienen de culturas notorias por el autosacrificio y el pensamiento comunitario tienen un sentido de autoestima y aprecian ser tratadas como individuos.
- Toda persona aprecia la consideración de que su opinión es importante para usted.

En el cuadro 8.14 se resumen las ventajas y problemas de las entrevistas.

Tasas de respuesta

La *tasa de respuesta* es el número de personas que participa realmente en una encuesta dividido entre el número total de personas invitadas a participar, en porcentaje. Una buena evaluación proporciona el número de personas (o unidades, por ejemplo, orga-

Recuadro 8.10 *Consejos para realizar entrevistas*

Los consejos siguientes pueden ayudar a los evaluadores a realizar entrevistas:

1. Permitir que los entrevistados sepan:
   - Cuál es el fin y el momento del estudio.
   - Por qué se los entrevista.
   - Cómo fueron seleccionados.
   - Cómo se utilizarán los datos.
   - Si la entrevista es confidencial.
   - Cuánto tiempo durará la entrevista.
   - Cuál es el contenido de la entrevista (compartiendo las preguntas por adelantado).
   - Si se desea hablar nuevamente con ellos.
   - Si recibirán copia del informe final.
   - Si se pondrá a su disposición un resumen de las notas del entrevistador, si lo desean.
2. Escoger un lugar y un tiempo tranquilos y libres de distracciones.
3. Idealmente, contar con una segunda persona para tomar notas.
4. Considerar la posibilidad de grabación de la entrevista. Si se hace, verificar con el entrevistado y obtener su permiso antes de grabar.
5. Atenerse al guión. Si se hacen preguntas cerradas, preguntar exactamente en la forma como fueron escritas. Si se hacen preguntas abiertas, "seguir el flujo" en lugar de dirigirlas siempre.
6. Estar consciente de las normas culturales, como el contacto visual, las preguntas directas y los problemas de género.
7. Mantener el equilibrio: si se pregunta, por ejemplo, a los entrevistados sobre lo que piensan son los soportes principales de un proyecto, seguir con lo que piensan que son las principales barreras.
8. Evitar hacer preguntas de tipo "por qué" si se considera como agresivo o crítico.
9. Aceptar lo que los entrevistados digan con empatía y sin juzgar.
10. Tomar buenas notas sin distraerse de la conversación.
    - Mantener el contacto visual mientras se escribe.
    - Escribir palabras o frases clave, no un recuento verbal de la entrevista.
    - Si el entrevistado dice algo que se desea capturar, solicitar a la persona repetirlo o terminar lo que se está escribiendo antes de hacer la pregunta siguiente.
    - Si el entrevistado dice algo que pueda citarse, preguntar si se pueden utilizar sus palabras exactas.
11. Escribir la entrevista.
    - Tener presente que cada palabra y cada idea son potencialmente valiosas.
    - Tomar el tiempo para escribir las notas con el mayor cuidado y profundidad posibles.
    - Efectuar por lo menos una limpieza breve de las notas inmediatamente después de la entrevista (dejar tiempo suficiente entre las entrevistas).
    - Escribir notas completas dentro del término de un día después de la entrevista.
12. En situaciones delicadas, considerar compartir las notas con el encuestado para verificar la recolección de respuestas y estar de acuerdo con lo que se ha dicho.

Fuente: los autores.

Recuadro 8.11 *Consejos para realizar entrevistas en diferentes culturas*

Los consejos siguientes pueden ayudar a los evaluadores a realizar entrevistas en entornos de desarrollo:

1. Antes de cada entrevista, aprender más sobre la cultura de la persona que se va a entrevistar.
2. Discutir técnicas de entrevista con alguien familiarizado con la cultura, o utilizar un entrevistador de la misma cultura que el entrevistado.
3. Averiguar tanto como sea posible sobre características culturales, por ejemplo:

    - Cantidad de espacio físico que debe haber entre el entrevistador y la persona entrevistada.
    - Cantidad apropiada de contacto visual.
    - Significado de inflexiones de voz, movimiento de la cabeza y demás lenguaje corporal durante una conversación.
    - Forma apropiada de vestir.
    - Forma apropiada de saludo.
    - Papeles de géneros.

Los asuntos de género son particularmente importantes. Por ejemplo, en algunas culturas puede ser inapropiado que un entrevistador masculino entreviste en una habitación a una mujer a solas. En ciertas culturas los hombres pueden cerrarse como una ostra si los va a entrevistar una mujer. Utilizar entrevistadores del mismo sexo o contar con la presencia de un testigo del sexo opuesto constituye un procedimiento necesario en dichas situaciones.

Cuadro 8.14 *Ventajas y problemas de las entrevistas*

Ventajas	• Pueden ser estructuradas, semiestructuradas, no estructuradas o una combinación de ellas, en general son semiestructuradas.
	• Pueden explorar profundamente temas complejos.
	• Son tolerantes con los errores: las preguntas no claras pueden clarificarse durante la entrevista y modificarse para entrevistas posteriores.
	• Pueden durar una hora o más, dependiendo de la importancia y el interés.
	• Ofrecen al evaluador un sentido intuitivo de la situación.
Problemas	• Pueden ser costosas, intensivas en mano de obra y requerir mucho tiempo.
	• No indican por qué distintas personas tienen distintos puntos de vista.
	• Como resultado de la audiencia selectiva por parte del entrevistador, pueden omitir información no conforme a las creencias preexistentes.
	• Requieren sensibilidad cultural, particularmente en relación con temas de género.

Fuente: los autores.

nizaciones) invitadas a la encuesta; el número de encuestados; la tasa de respuesta y una descripción de los esfuerzos realizados para incrementar la tasa de respuesta (llamadas telefónicas de seguimiento, cartas).

Tasa de respuesta: Número de personas que participa en una encuesta, dividido entre el número total de personas invitadas a participar, en porcentaje.

Las bajas tasas de respuesta sugieren una muestra sesgada (autoseleccionada), lo cual es un problema porque quienes eligen participar pueden ser distintos de quienes eligen no participar. Supóngase, por ejemplo, que una organización realiza una encuesta sobre actitudes para todos los empleados, pero solamente responde un 30%. Si los empleados más insatisfechos tendieron a llenar la encuesta mientras los satisfechos no, los datos no revelarán la opinión de todos los tipos de empleados. Sería un error tomar decisiones basadas en estos resultados sin recopilar más información sobre quiénes respondieron y quiénes no. Las encuestas con tasas de respuesta inferiores a un 70% no deben reportarse, a menos que el evaluador haya analizado el sesgo de respuesta y demostrado la validez de la encuesta (Jackson y Agrawal, 2007).

Un evaluador puede analizar la demografía de los encuestados para verificar si se empareja con la población general en términos de características como edad y género. Si es así, el evaluador puede proceder, pero sólo con una calificación apropiada de los resultados. En algunos casos, el evaluador puede hacer seguimiento directamente a algunos encuestados para establecer si existen patrones. Si la demografía de los que respondieron y los que no respondieron es distinta, deben calificarse apropiadamente los resultados de la encuesta. Los evaluadores deben informar dichos análisis de quienes no respondieron y cómo afectan potencialmente los hallazgos.

Los resultados de encuestas con bajas tasas de respuesta deben informarse siempre en cuanto al número de encuestados, reconociendo que estos resultados pueden reflejar o no con precisión al grupo general. Esto es, deben expresarse los datos como "de los 87 encuestados" o "un 75% ($N = 60$) de los encuestados respondió…". No deben utilizarse porcentajes si el total de los que enviaron la encuesta es menor de 50. Una proporción como "75% de los encuestados", cuando se aplica a un total de 45 encuestados es bastante diferente de un 75% de 3.000 encuestados.

Herramienta 5: grupo de discusión

Un *grupo de discusión* es una metodología de evaluación cuantitativa en la cual se reúnen pequeños grupos de personas para discutir temas específicos con la guía de un moderador. La estructura del grupo de discusión puede aparecer como informal, pero, en realidad, utiliza un guión que incluye un conjunto de preguntas abiertas. Un moderador capacitado dirige las discusiones e indaga o hace preguntas adicionales según la situación lo amerite.

Grupo de discusión: Metodología de evaluación cuantitativa en la cual se reúnen pequeños grupos de personas para discutir temas específicos con la guía de un moderador.

El proceso de grupo tiende a producir más información que las entrevistas individuales, pues las personas expresan distintos puntos de vista y dialogan entre ellas. Pero un

grupo de discusión no es simplemente una entrevista de grupo. El moderador facilita el diálogo y explora las razones y sentimientos de los integrantes. La conversación con mucha frecuencia es no lineal y los participantes traen a colación información o perspectivas diferentes en cualquier momento.

Los grupos de discusión se utilizan para recopilar datos sobre:

- Interacción de grupo.
- Complejidad de recursos.
- "Cómo" y "por qué" en lugar de "si" y "cuánto".
- Respuestas contextuales, no respuestas de tipo sí o no.
- Retroalimentación inmediata.
- Complejidad de comportamiento y motivaciones.
- Intensidad de opiniones.
- Opiniones sobre temas sensibles (Billson y London, 2004).

Pueden utilizarse para triangular métodos, y son también útiles cuando los encuestados no están cómodos con métodos de "papel y lápiz".

Los grupos de discusión no deben utilizarse para recopilar datos estadísticos y son también inapropiados donde:

- Las barreras idiomáticas son insuperables.
- El evaluador tiene poco control sobre el grupo.
- No puede establecerse la confianza.
- No puede garantizarse una libre expresión.
- La confidencialidad es crítica (Billson, 2004).

Un grupo de discusión incluye generalmente de 6 a 12 personas. La composición del grupo depende de su propósito. La mayoría de los grupos de discusión son homogéneos que incluyen personas de condición semejante (maestros en un grupo, estudiantes en otro; supervisores en un grupo, empleados en otro). Quién deba estar o no en el mismo grupo depende de la situación y de las culturas implicadas. La selección de los participantes no debe sesgar los resultados de la evaluación en ninguna forma predecible. No debe pasarse por alto ninguna categoría de participantes potenciales si se espera que sus respuestas alteren o expandan los resultados. En este caso de grupos adicionales potenciales de pertinencia, debería configurarse otro grupo de discusión.

Los grupos de discusión deben suministrar:

- Alrededores agradables y seguros (esencial).
- Refrescos (esencial).
- Incentivos monetarios (opcional).
- Transporte o arreglos para el cuidado de niños (opcional).
- Un moderador diestro (esencial).
- Un asistente que tome de notas (opcional).

Las sesiones se graban y se prepara, idealmente, una transcripción verbal para cada sesión. El grupo de discusión se inicia con una explicación clara de:

- Cuál es el propósito del grupo de discusión.
- Por qué es importante la opinión de los participantes.
- Cómo se seleccionan los participantes.
- Qué es un grupo de discusión.
- Cuáles son las reglas del proceso, incluyendo la regla básica: lo que se dice en el salón se queda en el salón.

El moderador guía los procesos, mantiene concentrado al grupo y se asegura de que todos los participantes tengan la oportunidad de expresar su opinión, y que unos pocos no dominen las conversaciones. El moderador hace un pequeño número de preguntas, utilizando una guía desarrollada específicamente para la sesión. Todas las preguntas son abiertas. La conversación progresa a partir de preguntas fáciles y conversacionales a preguntas más serias. Termina con preguntas de resumen y conclusiones que permiten la corrección de imprecisiones si es necesario y comentarios adicionales e ideas que se registran en grabaciones.

Los evaluadores son responsables de todos los aspectos del grupo de discusión, incluyendo:

- Quiénes deben estar en él.
- Cuántas sesiones deben realizarse.
- Dónde se realizan las sesiones.
- Cuándo se realizan.
- Qué preguntar a los participantes.
- Cómo analizar y presentar los datos.

Un grupo de discusión usual requiere seis a ocho semanas de tiempo de preparación. El tiempo depende de factores logísticos, la complejidad o la urgencia del proyecto, la accesibilidad de quienes toman las decisiones y la facilidad con que pueda convocarse a los participantes.

La evaluación del grupo de discusión funciona mejor si se realiza de la manera siguiente (Billson, 2004):

- *Paso 1:* clarificar las preguntas clave de la evaluación. Si los clientes y evaluadores no son claros con respecto a las preguntas clave que se supone serán respondidas en los grupos de discusión, todo el proceso será frustrante.
- *Paso 2:* diseñar el enfoque de la evaluación. Utilizar el fin de la evaluación y las preguntas clave, diseñar el enfoque general y el flujo de los temas para llegar a las preguntas clave.
- *Paso 3:* desarrollar la guía del moderador, que debe incluir todo el protocolo necesario (estructura o código de comportamiento) para la evaluación en forma que no sesgue las respuestas y dirija al grupo hacia los temas clave.

- *Paso 4:* convocar a los participantes.
- *Paso 5:* especificar las técnicas de la moderación. Identificar y utilizar buenas técnicas de moderación durante las sesiones del grupo de discusión.
- *Paso 6:* interrogar a los observadores, evaluadores y clientes y registrar información adicional. Inmediatamente después de cada grupo de discusión cuando los datos están frescos, compartir las percepciones generadas durante el grupo de discusión con los clientes y otras partes interesadas. Registrar información adicional no discutida abiertamente (impresiones, conclusiones, etcétera) para usarla en el paso siguiente.
- *Paso 7:* analizar los datos. Si ha funcionado bien el grupo de discusión, producirá una montaña de palabras e ideas. Estos datos cualitativos requieren técnicas analíticas especiales, en particular análisis de contenido (*ver* capítulo 10).
- *Paso 8:* presentar los hallazgos. Informar los hallazgos de forma que sean significativos y útiles para otros, en particular el cliente. Utilizar formatos orales, escritos o de video, o una combinación de cualquiera de ellos.

Las sesiones del grupo de discusión duran normalmente de una a dos horas. Para la mayoría de los proyectos, 100 minutos es tiempo suficiente para proseguir hasta el final de cada línea principal de cuestionamientos sin agotar o aburrir a los participantes o al moderador (Billson, 2004). Los grupos de discusión no deberían sobreprogramarse. Dos o tres sesiones de grupo de discusión en un día constituyen lo máximo que un moderador puede manejar.

Los grupos de discusión deberían realizarse, de ser posible, en un entorno neutral. La ubicación debe ser de fácil acceso. Usualmente, los participantes se sientan alrededor de una mesa, o en sillas arregladas en forma de círculo para facilitar la conversación entre ellos. Es posible que estén presentes algunos observadores en el salón mientras se realiza el grupo de discusión. Se les debería presentar al principio de la sesión y explicarse la razón de su presencia.

Las grabadoras de audio y video son extremamente útiles al proporcionar una transcripción precisa de una sesión. La mayoría de los participantes olvidan pronto que se les está grabando. Si no se graba la sesión, deberían estar presentes dos personas para tomar notas a fin de documentar las sesiones. La toma de notas en un computador portátil puede acelerar el análisis de los datos iniciales y la escritura de informes. El moderador no debería tomar notas.

No existe una regla fija respecto a cuántas sesiones de grupo de discusión realizar. La regla general es realizarlas hasta que surjan los mismos temas o no surja nueva información, lo cual sucede usualmente después de tres a seis sesiones. Puede ser útil hacer una serie central de preguntas de cada grupo de discusión, y luego agregar diferentes preguntas o indagaciones más extensas una vez esté claro que los temas son consistentes. Los evaluadores deben interrogarse unos a otros después de cada sesión, de modo que puedan tomarse decisiones sobre ajustes a los protocolos en sesiones posteriores.

Las actividades diarias de los participantes deben tenerse en cuenta: los evaluadores deben ser sensibles a las demandas sobre el tiempo de los participantes, incluyendo el tiem-

po de transporte de ida y vuelta; la participación en una sesión puede tomar un día o más. Un grupo de discusión programado al final de la tarde puede interferir con la preparación de una cena en la noche, lo que limitaría el número de personas dispuestas a participar.

Los grupos de discusión requieren considerable tiempo de preparación y plazo de realización. En el recuadro 8.12 se presentan consejos sobre la escritura de preguntas para grupos de discusión.

La sesión del grupo de discusión puede dividirse en cuatro fases distintas. Éstas y los pasos de cada una se presentan en el cuadro 8.15.

El propósito de las preguntas de grupos de discusión del recuadro 8.13 es ilustrar los tipos de preguntas que el evaluador puede desear hacer al utilizar esta herramienta.

En el cuadro 8.16 se resumen las ventajas y problemas de los grupos de discusión.

Herramienta 6: diarios, bitácoras y listas de revisión propias

Otro método de recopilar datos es utilizar diarios o listas de revisión propias.

Diarios o bitácoras

Un diario (o bitácora) es un informe propio escrito por un participante, usualmente con mantenimiento cada día. Los diarios se utilizan para capturar información detallada

Recuadro 8.12 *Consejos para diseñar preguntas de grupos de discusión*

Los consejos siguientes pueden ayudar a los evaluadores a obtener más de las sesiones de grupos de discusión:

- Piense primero sobre cómo se utilizará la información recabada del grupo de discusión. Asegúrese de que todas las preguntas lleven, directa o indirectamente, a información utilizable.
- Evite una redacción vaga y confusa.
- Haga una pregunta a la vez.
- Haga preguntas abiertas.
- Evite supuestos orientadores (conducentes) o engañosos.
- Evite preguntas que introduzcan sesgo y tergiversen las respuestas.
- Evite suministrar respuestas alternativas.
- Hágalo interesante. Las preguntas mecánicas producen respuestas mecánicas.
- Utilice el "embudo de protocolo", que empieza con preguntas sobre temas conceptuales generales y termina con preguntas inquisitivas.

El embudo de protocolo

Fuente: Billson, 2004: 29.

Cuadro 8.15 *Cuatro fases de acciones de grupo de discusión y moderador*

Fase	*Acciones*
I. Preámbulo o declaración de apertura	El moderador tranquiliza a los participantes, empezando con preguntas para romper el hielo.
	El moderador explica el propósito del grupo de discusión.
	El moderador presenta las reglas de terreno.
	Cada uno de los asistentes se presenta.
II. Presentaciones y calentamiento	Los participantes relatan experiencias y papeles sobre el tema.
	El moderador estimula la interacción del grupo y el pensamiento sobre el tema.
	El moderador inicia con las preguntas más inocentes y sencillas.
III. Cuerpo principal de discusión del grupo	El moderador pasa a preguntas más conminatorias o sensibles y complejas.
	El moderador produce respuestas profundas.
	Los datos emergentes se conectan para crear una base compleja e integrada para el análisis. Se asegura la participación general.
	El moderador termina con las preguntas cerradas.
IV. Clausura	El moderador resume y refina los temas clave.
	El moderador presenta teorías, impresiones y corazonadas a los integrantes del grupo para su reacción, confirmación o clarificación y corrección.
	El moderador invita a los participantes a una ronda de comentarios o percepciones finales ("enseñanzas clave adquiridas"), más "últimos pensamientos", cosas omitidas y cualquier otra cosa que los participantes deseen que conozca el equipo de evaluación.
	El moderador agradece a los participantes y distribuye el incentivo que se haya ofrecido.

Fuente: los autores.

Recuadro 8.13 *Muestra de preguntas de grupo de discusión*

El guión siguiente puede ser útil para modelar preguntas durante un grupo de discusión.

Presentaciones: empecemos pidiéndole a cada uno de ustedes que se presente. Por favor, mencione su título y en dónde trabaja, la naturaleza de su experiencia en la recepción de servicios de este programa de clínicas de salud, incluyendo por qué fue allí.

Sobre percepciones: los críticos del programa gratis de clínicas de salud sostienen que presta un servicio de baja calidad. ¿Cuál fue su experiencia en la recepción de servicios de las clínicas gratis?

Sobre gobernanza: encuestas recientes sugieren que los presupuestos actuales para los distritos no llegan a las clínicas locales. ¿Cuál es su evaluación de estos problemas basado en su experiencia de recepción de servicios de las clínicas?

Fuente: adaptado de Billson, 2004.

Cuadro 8.16 *Ventajas y problemas de los grupos de discusión*

Ventajas	Pueden realizarse con relativa rapidez y facilidad.
	Toman menos tiempo del personal que las entrevistas profundas y personales.
	Permiten flexibilidad para hacer cambios en procesos y preguntas.
	Pueden explorar distintas perspectivas.
	Pueden ser divertidos para los participantes.
Problemas	El análisis consume mucho tiempo.
	Los participantes pueden no ser representativos de la población, posiblemente sesgando los datos.
	El grupo puede estar influido por el moderador o por los miembros del grupo predominante.

Fuente: los autores.

sobre eventos de la vida diaria de las personas. Pueden utilizarse para estudiar fenómenos como redes sociales, comportamiento de salud, enfermedad y asociado, dieta y nutrición, trabajo en parcelas, hábitos de estudio, uso de anticonceptivos, prácticas de crianza de niños. Además de captar las actividades de interés en un tiempo particular, los diarios pueden incluir alguna caracterización de las actividades. Por ejemplo, a los fumadores inscritos en un programa para dejar de fumar se les puede solicitar utilizar un diario para registrar cuándo sienten una urgencia fuerte de fumar, dónde estaban y qué hacían en ese momento. La idea es que si los patrones son conocidos, los participantes pueden ejercer alguna acción sobre ellos. Si el final de una comida es un activador para fumar, pueden romper el patrón abandonando rápidamente la mesa después de comer.

Diario: Informe propio escrito por un participante, usualmente con mantenimiento cada día.

Los diarios se han utilizado en una escala nacional amplia y también en programas pequeños (Stafford, 2006). Los diarios constituyen herramientas útiles para complementar otras fuentes de datos (cuadro 8.17).

El conservar un diario requiere generalmente alfabetización y disposición de dedicar tiempo para mantenerlo. En algunos casos, sin embargo, los diarios pueden utilizar una lista de verificación para reducir la necesidad de alfabetización. Por ejemplo, un diario con imágenes se ha utilizado para recopilar datos sobre el consumo y el gasto en hogares en Tanzania y Gambia (Wiseman, Conteh y Matovu, 2005).

¿Qué cantidad de tiempo es correcta para utilizar una lista de revisión? En general, es necesario por lo menos una semana para obtener una muestra suficientemente grande de comportamientos, y más de un mes sería demasiado. Después de eso, el mantenimiento del diario se vuelve tedioso y disminuye la anotación de registros.

Usualmente, los participantes reciben cuadernillos con instrucciones claras y un ejemplo de una entrada en el diario completa. Los términos no usuales se explican, y en la última página del cuadernillo se pregunta si el diario cubre un período usual o no, y también para que los participantes compartan cualquier comentario no incluido en las entradas del diario.

001El camino.indd 321 08/11/2010 03:27:54 p.m.

Cuadro 8.17 Lineamientos para utilizar diarios o bitácoras

Acción	*Detalles*
Convocar personalmente a los participantes	Estimular la participación, empleando entrevistadores altamente motivados y personales.
	Apelar al altruismo y al sentido de servicio.
	Asegurar la confidencialidad.
	Ofrecer incentivos por anotaciones completas.
Proporcionar un cuadernillo a cada participante	Empezar con instrucciones claras sobre cómo llevar un diario.
	Incluir ejemplos de entradas de diario completas.
	Incluir prácticas de memoria cortas.
	Explicar todos los términos, por ejemplo, *evento* o *sesión*.
	En la última página, preguntar si el período cubierto por el diario es el usual o no para algunos comentarios o clarificaciones.
	Incluir un calendario, en donde se indique el vencimiento para una entrada.
Tener en cuenta el tiempo para la recolección de datos	Si el período es muy largo, puede ser fatigoso o tedioso para los participantes.
	Si es muy corto, en el diario podría faltar un comportamiento o un evento.

Fuente: los autores.

Lista de control llenada por los mismos entrevistados

Es un cruce entre un cuestionario y un diario. Se solicita a los participantes seguir la pista a un conjunto específico de actividades o de eventos, que se listan de modo que los encuestados puedan marcarlas fácilmente. Las listas de verificación pueden llenarse diaria o semanalmente, o cada vez que ocurra un evento o actividad particular. Por ejemplo, podría utilizarse una lista de verificación para capturar las horas en que se deben tomar píldoras contra la malaria, y anotar el número y tiempo de viajes a una conexión de agua.

Lista de control: Conjunto específico de actividades o de eventos que los encuestados marcan como aplicables.

Las listas de verificación propias se utilizaron en un estudio de parteras en Indonesia (Abadía, 2002). El proyecto se basó en la teoría de que los proveedores necesitan refuerzos después de capacitar para el cambio de comportamiento. El objetivo fue mejorar la calidad de la asesoría de las parteras en planificación familiar. Las parteras participantes completaron una semana de autoevaluación en destrezas específicas de comunicación interpersonal para aumentar su conocimiento de sus propios patrones de comunicación, analizarlos y tomar medidas para mejorarlos.

Una lista de verificación propia es más fácil de completar y analizar que un diario. Sin embargo, crear una lista de verificación requiere que el evaluador comprenda la situación de manera muy específica.

En el cuadro 8.18 se resumen las ventajas y problemas de los diarios y listas de verificación.

Herramienta 7: juicio de expertos

Juicio de expertos: Opinión de un experto o de un panel de expertos.

Algunas veces tiene sentido vincular a expertos como fuente de información o de opinión. Considérese el papel de un crítico literario o de cine. La gente utiliza su juicio para informar sus propias decisiones o elecciones. Las fuerzas especiales del gobierno constituyen una forma de juicio de expertos.

El *juicio de expertos* puede obtenerse de un solo experto o de un panel de expertos. Un panel de expertos puede llevar a cabo una revisión de un hospital como parte del proceso de acreditación. Este grupo de expertos visita el hospital para investigar sus programas, instalaciones y personal. Un experto individual (por ejemplo, un inspector escolar) puede visitar colegios para evaluar la sede, la administración y los profesores.

Mediante el juicio de expertos pueden utilizarse criterios y procedimientos profesionales preestablecidos, o éstos pueden ser para fines específicos. Con criterios preestablecidos, éstos suelen ser estándares publicados e instrumentos que aseguran que los expertos hagan las preguntas y evalúen de una manera consistente. Por ejemplo, una agencia de financiación desea evaluar la calidad de un programa de capacitación para el que se estableció un plan de estudios, con metas, objetivos y efectos directos esperados. Un experto totalmente versado en el plan de estudios, las metas, los objetivos y efectos directos examina el programa de capacitación y basa la evaluación en la documentación.

Las revisiones con fines específicos se efectúan mediante programaciones irregulares, según lo indiquen las circunstancias y usualmente se realizan para un propósito específico. Tales revisiones no tienen usualmente estándares predeterminados para uso

Cuadro 8.18 *Ventajas y problemas de los diarios y las listas de verificación*

Ventajas	Pueden capturar datos profundos y detallados ("datos ricos") que de otra forma pueden quedar olvidados.
	Pueden recopilar datos sobre el uso del tiempo de las personas.
	Pueden recopilar información sensible.
	Pueden complementar entrevistas y suministrar datos más ricos.
	Requieren alguna alfabetización.
Problemas	Pueden cambiar el comportamiento, porque las personas saben que se observa su comportamiento.
	Requieren compromiso y autodisciplina, registro honesto y preciso de información por parte del participante que no puede recordar comportamientos con precisión.
	Los datos registrados pueden estar incompletos, o el participante puede haber esperado un tiempo para registrar la información y no recuerda correctamente ("diarios de mañana").
	Puede ser difícil leer la caligrafía de las personas.
	Pueden ser de difícil comprensión las frases del diario.

Fuente: los autores.

de los expertos, quienes establecen los estándares para emitir su juicio. Un ejemplo de una revisión con fines específicos es el de un director de un programa de tratamiento de aguas que visita el sitio para tener una idea de cómo prosigue el programa.

La selección de expertos debe pasar la prueba de la persona razonable: ¿Pensaría una persona razonable que el juicio de este experto o grupo de expertos es confiable? Los expertos deben reflejar un conjunto diverso de puntos de vista, experiencias y papeles.

Es importante establecer criterios para seleccionar expertos basados no sólo en el reconocimiento de las personas como expertos, sino también en los criterios siguientes:

- Áreas de profesionalismo
- Diversidad de perspectivas
- Diversidad de opiniones políticas
- Diversidad de pericias técnicas

El fundamento para la selección de los expertos debe indicarse en el informe de evaluación.

El Internacional Council on Security and Development (2006) utilizó un panel de expertos para revisar la eficacia de la International Narcotics Control Board. En este caso, fue particularmente importante no sólo contar con expertos reconocidos, sino también garantizar que fuesen seleccionados de diferentes partes del mundo, especialmente de aquellas con grandes problemas de narcóticos reconocidos.

Mientras un panel de expertos no es considerado como un enfoque de evaluación fuerte, en algunas circunstancias puede ser el mejor enfoque dadas las restricciones de tiempo y recursos. Los paneles de expertos son más apropiados en el diseño del programa o el proyecto, y al principio de las etapas intermedias de implementación que para las evaluaciones de impacto. Son especialmente útiles en las evaluaciones rápidas.

En el cuadro 8.19 se presentan las ventajas y problemas del juicio de expertos.

Herramienta 8: técnica Delphi

La *técnica Delphi* permite que los expertos que viven en distintas localidades dialoguen y lleguen a un consenso a través de un proceso iterativo. Se hacen preguntas específicas a los expertos y se envían sus respuestas a una fuente central, se resumen y se retroalimenta a los expertos, quienes hacen comentarios sobre el resumen. Tienen libertad para cuestionar puntos de vista particulares o agregar nuevas perspectivas, suministrando información adicional. Puesto que no se sabe lo que dijo cada quien, se evita el conflicto.

Cuadro 8.19 *Ventajas y problemas del juicio de expertos*

Ventajas	Rápido y relativamente económico.
Problemas	Débil para evaluaciones de impacto.
	Puede basarse mayormente en percepciones.
	El valor de los datos recopilados depende de la credibilidad que se tenga del grupo.

Fuente: los autores.

Técnica Delphi: Técnica que produce información y juicios de los participantes para facilitar la resolución de problemas, la planeación y la toma de decisiones sin enfrentar personalmente a los participantes.

La técnica Delphi produce información y juicios de los participantes para facilitar la resolución de problemas, la planeación y la toma de decisiones, sin enfrentar personalmente a los participantes (Dunham, 1996). Los expertos comparten información por correo, fax o correo electrónico. Este método puede ayudar a los grupos a lograr un consenso sobre problemas, estrategias y prioridades. Se ha utilizado para pronosticar el potencial turístico en Sudáfrica (Kaynak, Bloom y Leibold, 1994), lograr un consenso sobre las prioridades para políticas nacionales sobre fármacos en países en desarrollo (Rainhorn, Brudon-Jakobowicz y Reic, 1994), y establecer prioridades para erradicar la esquistosomiasis (fiebre de caracol) en Kenia (Kirigia, 1997).

La técnica Delphi requiere un coordinador, quien identifica los expertos como participantes, organiza las solicitudes de información, resume la información recibida y se comunica con los participantes. El trabajo del coordinador puede tomar mucho tiempo. Coordinar la técnica Delphi por correo electrónico con 20 participantes y tres rondas de cuestionarios podría tomar de 30 a 40 horas del tiempo del coordinador.

El proceso de la técnica Delphi incluye los pasos siguientes:

1. Identificar e invitar a los expertos a participar. Establecer los criterios necesarios de los expertos, usualmente un grupo con diversas pericias. Invitar a los expertos a unirse.
2. Identificar el problema y solicitar ideas. Por ejemplo, preguntar: "¿Qué medidas pueden tomarse para una respuesta más rápida a las consultas de los pacientes entre visitas? Preparar y enviar el primer cuestionario, que solicita a cada participante unirse a una sesión de intercambio de ideas individuales para generar tantas ideas como sea posible para tratar el problema.
3. Hacer que los participantes llenen el primer cuestionario. No es necesario que las ideas estén completamente desarrolladas y, en realidad, es preferible que cada idea se exprese en un solo enunciado o una frase, sin intención de evaluar o justificar estas ideas. Los participantes devuelven entonces la lista al coordinador.
4. Crear y enviar el segundo cuestionario. El coordinador prepara y envía un segundo cuestionario a los participantes, que contiene todas las ideas enviadas como respuesta al primer cuestionario. (Sólo el coordinador sabe con qué ideas contribuyó cada experto.) Se da espacio para que los participantes refinen cada idea, comenten sus fortalezas y debilidades, e identifiquen nuevas ideas.
5. Recopilar las respuestas al segundo cuestionario. Los participantes, en forma anónima, registran sus respuestas al segundo cuestionario y las devuelven al coordinador.
6. Crear y enviar el tercer cuestionario. El coordinador prepara y envía un tercer cuestionario, en el que se resumen los aportes del paso anterior y solicita clarificaciones, fortalezas y debilidades adicionales y nuevas ideas.
7. Continúa el proceso. Si se desea, el coordinador efectúa nuevas rondas del proceso hasta que ya no surjan nuevas ideas y se hayan identificado todas las fortalezas, debilidades y opiniones.

8. Lograr la solución, la que puede ocurrir en una de dos formas.

   - Si surge un consenso sobre las ideas predominantes que se han evaluado, se declara terminado el ejercicio. El producto final es una lista de ideas, con indicación de sus fortalezas y debilidades.
   - Si no hay consenso, el coordinador efectúa una evaluación formal de las opiniones del grupo sobre los méritos de las ideas. Hay varias maneras de hacer una evaluación formal. Un enfoque es que el coordinador prepare un cuestionario en el que se relacionen todas las ideas y solicite a los participantes clasificarlas una por una. Podría utilizarse una escala de siete puntos, de cero (ningún potencial para tratar el problema) a 7 (potencial muy alto para tratar el problema). De utilizarse este enfoque, los participantes envían los formularios de clasificación al coordinador, quien compila los resultados y ordena las ideas por rango con base en las evaluaciones. Un segundo enfoque es que el coordinador solicite a cada integrante identificar las cinco ideas principales y asigne cinco puntos a la más promisoria, cuatro a la siguiente, y así sucesivamente. Los votos se envían al coordinador, quien recuenta los resultados y prepara un informe en donde se anota el orden de clasificación de las ideas basado en el número total de puntos recibidos, e indica el número de personas que votó por cada idea (Dunham, 1996).

En el cuadro 8.20 se presentan las ventajas y problemas de la técnica Delphi.

Herramienta 9: otras herramientas de medición

Para algunos proyectos o programas, los evaluadores encontrarán otras herramientas de medición, por ejemplo, dispositivos electrónicos, químicos y mecánicos. Las herramientas como pruebas en papel y lápiz, y pruebas o evaluaciones de destrezas, pueden utilizarse para recopilar datos para algunos proyectos y programas. A continuación se presentan ejemplos de algunas de las otras herramientas utilizadas para la recolección de datos.

Muchos proyectos y programas (especialmente aquellos del sector de la salud) requieren la medición de la condición fisiológica para determinar los efectos de las intervenciones. Se utilizan escalas para recopilar datos sobre el peso de los sujetos. La altura, longitud y circunferencia de apéndices pueden medirse utilizando medidas de cinta, varas de medir u otros dispositivos. Los cronómetros pueden utilizarse para medir el tiempo de terminar una tarea. En algunos casos se utilizan herramientas especializadas para recopilar datos fisiológicos adicionales sobre colesterol, presión sanguínea, azúcar, composición corporal, capacidad pulmonar, capacidad aeróbica, fuerza muscular, flexibilidad de junturas y otras variables.

Las pruebas químicas de agua de pozo y suelos pueden utilizarse para medir la calidad del agua o del suelo. Con otras pruebas pueden medirse el punto de fusión, la presión del vapor y los factores de bioconcentración para determinar las propiedades físicas y

Cuadro 8.20 *Ventajas y problemas de la técnica Delphi*

Ventajas	Permite el anonimato de los participantes.
	No es costosa.
	Está libre de presión social, influencia de personalidades y predominio individual.
	Es conducente al pensamiento independiente y la formulación gradual.
	Permite compartir la información y el razonamiento entre los participantes.
Problemas	Puede no ser representativa
	Tiende a eliminar posiciones extremas y forzar a un consenso de mitad de camino.
	Requiere destreza en la comunicación escrita.
	Requiere tiempo y compromiso de los participantes.

Fuente: Michigan State University Extension, 1994.

químicas de cada químico específico. Las pruebas como estas contribuyen a identificar los riesgos ambientales y de la salud (U.S. EPA, 2007). Cuando se utilizan dispositivos de medición, es importante revisar la calibración de los instrumentos, y hacer las correcciones necesarias. En los proyectos y programas sobre nutrición se pueden utilizar evaluaciones de los componentes químicos o nutritivos de los alimentos.

Las pruebas de aptitud o los logros pueden utilizarse para evaluar el conocimiento y los hechos, y se utilizan comúnmente para programas de capacitación y educación, pero son aplicables también en muchos otros campos.

En el pasado, se administraban muchas pruebas utilizando papel y lápiz. Ahora pueden efectuarse electrónicamente, incluso por Internet.

Antes de escoger una prueba, es importante comprender la clase de prueba que mejor mide lo que el evaluador desea saber. Existen cuatro enfoques para las pruebas de logros:

- Referenciadas a normas.
- Referenciadas a criterios.
- Referenciadas a objetivos.
- Referenciadas a dominios (Fitzpatrick, Sanders y Worthen, 2004).

Las pruebas referenciadas a normas se utilizan primordialmente para comparar el desempeño de los estudiantes con el de otros que toman la misma prueba. Las pruebas referenciadas a normas se usan con mucha frecuencia para evaluar el progreso. Muchos distritos escolares utilizan pruebas referenciadas a normas para efectuar comparaciones de sus estudiantes con los de otros distritos escolares sobre conocimientos y destrezas comúnmente aceptadas. Una debilidad en las pruebas referenciadas a normas es la de identificar "conocimientos y destrezas comúnmente aceptadas". Las diferentes culturas, sociedades y entornos pueden tener objetivos de instrucción muy diferentes (Fitzpatrick, Sanders y Worthen, 2004).

Las pruebas referenciadas a criterios se utilizan para medir el desempeño frente a un criterio absoluto. Un país puede establecer criterios frente a los cuales juzgar el desempeño de los colegios o los distritos escolares (no estudiantes individuales), y usarlos para determinar cómo se comportan los colegios o los distritos. Los criterios se utilizan a menudo para evaluar un plan o programa de estudios (Fitzpatrick, Sanders y Worthen, 2004).

Las pruebas referenciadas a objetivos se basan en los objetivos o efectos directos identificados por un programa educativo o de capacitación. Los renglones para la prueba se basan en objetivos o efectos directos declarados. Las pruebas referenciadas a objetivos se utilizan para suministrar retroalimentación de evaluación formativa para ayudar a los maestros o instructores a examinar campos que cumplen los objetivos o efectos directos y los que necesitan mejoras (Fitzpatrick, Sanders y Worthen, 2004).

Las pruebas referenciadas a dominios se vinculan a un dominio de contenido que se mide, por ejemplo, con conocimientos de historia o matemáticas. El desarrollo de estas pruebas es costoso, pero ofrecen información valiosa, tal como "¿Cuánto saben nuestros estudiantes sobre geografía mundial?" o "¿Cuánto saben nuestros capacitados sobre género?" (Fitzpatrick, Sanders y Worthen, 2004).

Las pruebas referenciadas a normas y a criterios proporcionan estándares para juzgar el desempeño. Las pruebas referenciadas a objetivos y a dominios no proporcionan estándares. Proporcionan datos descriptivos sobre desempeño estudiantil sin juzgar estudiantes o colegios (Fitzpatrick, Sanders y Worthen, 2004).

En Internet están disponibles muchas pruebas estandarizadas (*ver* los vínculos al final de este capítulo) que pueden ayudar a evaluar la personalidad, el interés vocacional, las aptitudes, logros, desarrollo, inteligencia, lectura y destrezas de estudio.

Algunas partes interesadas desean conocer más sobre una intervención que los puntajes de los participantes en una prueba de conocimientos estandarizada. El desempeño en destrezas puede evaluarse utilizando dispositivos de simulación, desarrollo de portafolio, presentaciones orales y debates. Es importante emparejar la destreza requerida con la prueba que mejor demuestra el desempeño deseado.

Las pruebas de destreza son útiles para evaluar cualquier intervención que busque introducir nuevas destrezas (prácticas de sanidad a madres, prácticas agrícolas a agricultores, prácticas de mantenimiento de registros a funcionarios oficiales. En las evaluaciones para el desarrollo se intenta a menudo evaluar si se utilizó la nueva práctica, pero no puede evaluarse si se utilizó diestramente. Pueden utilizarse expertos capacitados en observar el desempeño para examinar y evaluar destrezas. Se establecen criterios y se califica a los sujetos su grado de cumplimiento de los criterios. La fotograma o videograma pueden utilizarse para documentar el desempeño.

Resumen

No existe una sola forma mejor de recopilar datos. De los muchos métodos descritos en el presente capítulo, el que deba utilizarse depende de las necesidades de la evaluación y las preguntas específicas que se traten.

Las evaluaciones deberían utilizar más de una técnica de recopilación de datos. Las decisiones sobre las técnicas de recopilación de datos que deben utilizarse deben basarse en:

- Lo que el evaluador necesita saber.
- Dónde se encuentran los datos.
- La disponibilidad de recursos y tiempo.
- La complejidad de los datos que se deben recopilar.
- La frecuencia de la recopilación de datos.

Actividades del capítulo 8

Ejercicio de aplicación 8.1: recopilar información de archivos

Se le pide utilizar los archivos de admisión para establecer las calificaciones y experiencia de estudiantes admitidos a un taller de capacitación en horticultura. Desarrolle un formulario corto de cinco preguntas que pueda utilizarse para recopilar los datos de estos archivos.

Ejercicio de aplicación 8.2: recopilar información de entrevistas

Se le ha solicitado desarrollar preguntas para una corta entrevista con el fin de evaluar reacciones de participantes a la calidad de un taller, con base en una conferencia a la que asistió recientemente. Desarrolle cinco preguntas abiertas que traten sobre el contenido, el nivel y la ejecución del taller. De ser posible, encuentre un colega que haya asistido al mismo taller o conferencia. Entreviste a su colega y permita que él lo entreviste, utilizando el instrumento desarrollado por usted. Luego escriba un resumen profundo de las entrevistas. Solicite a su colega criticarlo en cuanto a precisión, legibilidad y cobertura.

Ejercicio de aplicación 8.3: recopilar información de grupos de discusión

Se le ha solicitado diseñar un grupo de discusión para evaluar el impacto de una serie de talleres y un paquete de asistencia financiera para ayudar a mujeres a iniciar y operar sus propios pequeños negocios. Desarrolle un conjunto de cinco preguntas para un grupo de discusión que sea apropiado para hacerlas a mujeres que completaron el programa hace seis meses. Asegúrese de averiguar qué efectos había previsto el programa y cómo afectaron la vida, a los amigos y la familia de los participantes.

Referencias y lecturas adicionales

Abdallah, H. 2002. *Cost-Effectiveness of Self-Assessment and Peer Review in Improving Family Planning Provider-Client Communication in Indonesia.* Quality Assurance Project Case Study. Estudio realizado para la U.S. Agency for International Development por el Quality Assurance Project, Bethesda, MD.

Academy for Educational Development, Population Communication Services. 2002. "Session 10 Transect Walks and Observation". *Empowering Communities: Participatory Techniques for Community-Based Programme Development,* Vol. 2, *Participant Handbook.* Washington, DC: Academy for Educational Development. http://pcs.aed.org/manuals/cafs/handbook/sessions10-12.pdf.

BAsD (Banco Asiático de Desarrollo) y ADBI (Asian Development Bank Institute). 2004. *Improving Local Governance and Service Delivery: Citizen Report Card Learning Tool Kit.* http://www.citizenreportcard.com/index.html#.

Banco Mundial. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey". Documento de trabajo 126, Washington, DC.

———. 2006. "Citizen Report Cards: A Presentation on Methodology". Participation and Civic Engagement Group, Social Development Department, Washington, DC. http://info.worldbank.org/etools/docs/library/94360/Tanz_0603/Ta_0603/CitizenReportCardPresentation.pdf.

———. 2007. "Community-Driven Development". Washington, DC. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTCDD/0,,menuPK:430167~pagePK:149018~piPK:149093~theSitePK:430161,00.html.

BBC (British Broadcasting Corporation). 2008. "Data Validation and Verification". *GCSE Bitesize.* http://www.bbc.co.uk/schools/gcsebitesize/ict/databases/3datavalidationrev1.shtml.

Behling, Orlando y Kenneth S. Law. 2000. *Translating Questionnaires and Other Research Instruments: Problems and Solutions.* Quantitative Applications in the Social Sciences, Series 07-133. Thousand Oaks, CA: Sage Publications.

Billson, Janet Mancini. 2002. *The Power of Focus Groups for Social and Policy Research.* Barrington, RI: Skywood Press.

———. 2004. *The Power of Focus Groups: A Training Manual for Social, Policy, and Market Research: Focus on International Development.* Barrington, RI: Skywood Press.

Billson, Janet y N. T. London. 2004. "The Power of Focus Groups". International Program for Development Evaluation Training (IPDET); presentation, Ontario, julio.

Burgess, Thomas F. 2001. *Guide to the Design of Questionnaires,* ed. 1.1. University of Leeds, Reino Unido. http://www.leeds.ac.uk/iss/documentation/top/top2/top2-5.html.

Cloutier, Dorothea, Bill Lilley, Devon Phillips, Bill Weber y David Sanderson. 1987. *A Guide to Program Evaluation and Reporting.* University of Maine Cooperative Extension Service, Orono.

CNCSTE (Chinese National Centre for Science and Technology Evaluation), and IOB (Policy and Operations Evaluation Department). 2006. *Country-Led Joint Evaluation of the ORET/MILIEV Programme in China.* Amsterdam: Aksant Academic Publishers.

Cnossen, Christine. 1997. *Secondary Research.* Documento de aprendizaje 7, School of Public Administration and Law, Robert Gordon University, Aberdeen, Reino Unido.

Dawson, Susan y Lenor Manderson. 1993. *Methods for Social Research in Disease: A Manual for the Use of Focus Groups.* International Nutrition Foundation for Developing Countries, Boston. http://www.unu.edu/Unupress/food2/UIN03E/uin03e00.htm.

Denzin, K. 1978. *The Research Act.* Nueva York: McGraw-Hill.

Dunham, Randall B. 1996. *The Delphi Technique.* http://www.medsch.wisc.edu/adminmed/2002/orgbehav/delphi.pdf.

Early Childhood Research Institute on Measuring Growth and Development. 2008. *Early Communication Indicator (ECI).* http://cehd.umn.edu/ceed/projects/ecri/.

EuropeAid Co-operation Office. 2005. *Evaluation Methods.* ec.europa.eu/europeaid/evaluation/methodology/egeval/index_en.htm.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* Nueva York: Pearson.

Foddy, William. 1993. *Constructing Questions for Interviews and Questionnaires.* Nueva York: Cambridge University Press.

Fowler, Floyd J. Jr. 2002. *Applied Social Research Methods Series,* Vol. 1 *Survey Research Methods.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Frary, Robert B. 1996. "Hints for Designing Effective Questionnaires". *Practical Assessment, Research & Evaluation* 5 (3). http://PAREonline.net/getvn.asp?v=5&n=3.

Hofstede, G. 2001. *Culture's Consequences.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Jackson, Gregg y Rashmi Agrawal. 2007. "Guidelines for Developing Survey Instruments as a Part of the Designing and Conducting Surveys Workshop". International Program for Development Evaluation Training (IPDET); presentation, Ottawa, Ontario.

Kaynak, Erdener, Jonathan Bloom y Marius Leibold. 1994. "Using the Delphi Technique to Predict Future Tourism Potential". *Marketing Intelligence and Planning* 12 (7): 18-29. http://www.emeraldinsight.com/Insight/viewContentItem.do;jsessionid=372EF38765D6F8133CDF8DABFB2071B4?contentType=Article&hdAction=lnkpdf&contentId=854271.

Kirigia, J. M. 1997. *Economic Evaluation in Schistosomiasis Using the Delphi Technique to Assess Effectiveness.* http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9107365&dopt=Abstract.

Krueger, R. A. y M. A. Casey. 2000. *Focus Groups.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Kellogg Foundation. 2008. *"Tips on Using Tests and Assessments". Evaluation Toolkit.* http://www.wkkf.org/Default.aspx?tabid=90&CID=281&ItemID=2810034&NID=2820034&LanguageID=0.

Lofland, John. 1971. *Analyzing Social Settings.* Belmont, CA: Wadsworth.

Lofland, John y L. H. Lofland. 1995. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* 3ª ed. Belmont, CA: Wadsworth Publication.

McCaston, M. Katherine. 2005. *Tips for Collecting, Reviewing y Analyzing Secondary Data.* Partnership and Household Livelihood Security Unit. http://pqdl.care.org/pv_obj_cache/pv_obj_id_8F453F01C87B8BB24774628B95B42BBCBD020200.

McNamara, Carter. 2007. *General Guidelines for Conducting Interviews.* http://www.managementhelp.org/evaluatn/intrview.htm#anchor615874

Michigan State University Extension. 1994. *Delphi Technique.* http://web1.msue.msu.edu/msue/imp/modii/iii00006.html.

Miles, Matthew B. y A. Michael Huberman. 1994. *Qualitative Data Analysis an Expanded Sourcebook.* Thousand Oaks, CA: Sage Publications.

Minnesota Department of Health. 2007. *Community Engagement: Community Forums and Public Hearings.* http://www.health.state.mn.us/communityeng/needs/needs.html.

Narayan, Deepa. 1996. *Toward Participatory Research.* Documento técnico del Banco Mundial 307, Washington, DC. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1996/04/01/000009265_3980625172923/Rendered/PDF/multi0page.pdf.

O'Brien, G. 2001. *Data Flow, Application of Data Quality Flags y Data Validation Processes for CCAQS.* Propuesta preliminar. http://www.arb.ca.gov/airways/Documents/reference percent20tables/Marchpercent2006, percent202001/DataFlow_PrelimDraft.pdf.

Patton, Michael Q. 1987. *How to Use Qualitative Methods in Evaluation.* Thousand Oaks, CA: Sage Publications.

———. 2002. *Qualitative Evaluation and Research Methods.* 3ª ed. Thousand Oaks, CA. Sage Publications.

Paul, Samuel y Gopakumar K. Thampi. 2008. "Monitoring Public Service Delivery: Citizen Report Cards Score in India". *Capacity.org.* 35 (diciembre). PNUD, Nueva York. http://www.capacity.org/en/journal/tools_and_methods/citizen_report_cards_score_in_india.

Phuyal, Kamal. 2006. "Sharing Some Practical Experiences on Participatory Appreciative Planning Approach (PAPA): An Appreciative Approach to Working in Partnership with Community People". Documento presentado en el international workshop sobre "Action Learning: Lessons from the Field", organizado por la Faculty of Policy Studies, Chuo University, Tokio, octubre. http://www.scn.org/cmp/modules/emp-papa.htm.

Porteous, Nancy L., B. J. Sheldrick y P. J. Stewart. 1997. *Program Evaluation Tool Kit: A Blueprint for Public Health Management.* Ottawa: Ottawa-Carleton Health Department.

Rainhorn, J.-D. P. Brudon-Jakobowicz y M. R. Reich. 1994. "Priorities for Pharmaceutical Policies in Developing Countries: Results of a Delphi Survey". Boletín de la Organización Mundial de la Salud 72(2): 257-64. 72 (2): 257-64. http://74.125.47.132/search?q=cache:D1zbVUz7oG0J:whqlibdoc.who.int/bulletin/1994/Vol72-No2/bulletin_1994_72(2)_257-264.pdf+Use+of+the+delphi+technique+in+developing+countries&cd=3&hl=en&ct=clnk&gl=us&client=firefox-a.

Sanders, J. R. 2000. *Evaluating School Programs.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Stafford, Frank P. 2006. *Timeline Data Collection and Analysis: Time Diary and Event History Calendar Methods.* Department of Economics, University of Michigan, Ann Arbor. http://www.atususers.umd.edu/wip2/papers_i2007/Stafford_Diaries.pdf.

Stake, Robert E. 1995. *The Art of Case Study Research.* Thousand Oaks, CA: Sage Publications.

StatPac. 2007. *Survey Software, Question Wording.* http://www.statpac.com/surveys/index.htm#toc.

Sue, Valerie M. y Lois A. Ritter. 2007. *Conducting Online Surveys.* Thousand Oaks, CA: Sage Publications.

TC Evaluation Center. 2007. *Wording Questions: Some General Guidelines.* University of California, Davis. http://ucce.ucdavis.edu/files/filelibrary/5715/27621.pdf.

Trochim, William M. K. 2006. *Types of Data, Research Methods Knowledge Base.* http://www.socialresearchmethods.net/kb/datatype.php.

UNESCAP (United Nations Economic and Social Commission for Asia and the Pacific). 1999. *Guidelines on the Application of New Technology to Population Data Collection and Capture.* http://www.unescap.org/stat/pop-it/pop-guide/capture_ch04.pdf.

University of Wisconsin, Cooperative Extension. 1996. *Program Development and Evaluation, Collecting Evaluation Data: Direct Observation.* http://learningstore.uwex.edu/pdf/G3658-5.pdf.

U.S. EPA (U.S. Environmental Protection Agency). 2007. *CTSA-Chapter 2: Data Collection.* http://www.epa.gov/dfe/pubs/lithography/ctsa/ch02/ch02.html.

Wadsworth, Y. 1997a. *Do It Yourself Social Research.* 2ª ed. St. Leonards, Nueva Gales del Sur, Australia: Allen and Unwin.

———. 1997b. *Everyday Evaluation on the Run.* St. Leonards, Nueva Gales del Sur, Australia: Allen and Unwin.

Wengraf, Tom. 2001. *Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods.* Thousand Oaks, CA: Sage Publications.

Wiseman, V., L. Conteh y F. Matovu. 2005. "Using Diaries to Collect Data in Resource-Poor Settings: Questions on Design and Implementation". Health Policy and Planning 20 (6): 394-404. http://heapol.oxfordjournals.org/cgi/reprint/20/6/394.

Yin, Robert K. 2003. *Case Study Research: Design and Methods.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Sitios web

Alphabetical List of Aptitude or Achievement Tests. http://www.yorku.ca/psycentr/tests/alpha.html.
Evaluation Portal. http://www.evaluation.lars-balzer.name/.
International Council on Security and Development. 2006. "Think tank announces expert panel to review effectiveness of International Narcotics Control Board". http://www.icosgroup.net/modules/press_releases/narcotics_control_board.
Measurement Group. *Evaluation/Research Tools.* http://www.themeasurementgroup.com/evalbttn.htm.
Nielsen, J. 1997. *The Use and Misuse of Focus Groups.* http://www.useit.com/papers/focusgroups.html.
Standardized Monitoring and Assessment of Relief and Transitions. *Instruments and Tools.* http://www.smartindicators.org/IT.php?page=80.

001El camino.indd 334 08/11/2010 03:27:54 p.m.

CAPÍTULO 9

Seleccionar la estrategia de muestreo

En este capítulo se discute cómo determinar la cantidad de datos que se ha de recopilar y cómo seleccionar las fuentes de datos de modo que reflejen aproximadamente la población, y contribuyan a responder las preguntas de la evaluación.

Introducción al muestreo

Al haber decidido los evaluadores recopilar los datos de una *población* particular (de personas, clínicas de salud o de otras unidades), necesitan decidir si recopilar los datos de todas las unidades de interés o de un subconjunto de esas unidades. La recopilación de datos de todas las unidades se denomina *censo*, y *muestra*, la de un subconjunto de todas las unidades. En general, se prefiere un censo cuando la población es pequeña y el costo de la recopilación de datos es bajo. Cuando la población es grande o el costo de recopilar los datos es alto, usualmente se necesita recopilar datos de una muestra. También se utiliza el muestreo algunas veces, para minimizar la carga de los encuestados.

Población: Conjunto total de unidades sobre las cuales el evaluador desea hacer inferencias.

Censo: Colección de datos de una población entera.

Muestra: Subconjunto de una población sobre la que se recopilan datos.

Las personas utilizan muestras todo el tiempo. Por ejemplo, cuando se les hace una prueba de sangre para

verificar su salud, el laboratorio toma una muestra y no toda su sangre. Las pruebas se hacen utilizando esa muestra, y se supone que aquello que se encuentra en la muestra refleja con precisión lo que se encuentra en toda la sangre.

El muestreo no es sólo algo que se aplique a grandes estudios cuantitativos. Por ejemplo, aun cuando se haga una visita de campo altamente cualitativa durante una semana para evaluar un programa que se distribuye sobre una gran región geográfica, todavía los evaluadores deben pensar cuidadosamente sobre qué zonas de esa región deben investigar. Pueden introducirse sesgos si los funcionarios del programa, que desean mostrar lo mejor, seleccionan los sitios y los participantes que se van a estudiar. Esos sesgos pueden evitarse con una muestra seleccionada al azar. El conocimiento de los conceptos básicos del muestreo puede mejorar el grado en que una evaluación refleje con precisión lo que en realidad está sucediendo en un programa.

Tipos de muestras: aleatorias y no aleatorias

Cuando no se pueden recopilar los datos de cada país, cada persona o cada parcela, los evaluadores utilizan un subconjunto, o muestra, para el estudio. Las muestras pueden ser aleatorias o no aleatorias.

Muestreo aleatorio

Las *muestras aleatorias* son aquelllas en las que cada unidad de la población tiene igual probabilidad de ser seleccionada. Una lotería es un ejemplo de una muestra aleatoria, porque cada número tiene igual probabilidad de ser seleccionado como el número ganador.

Una ventaja del muestreo aleatorio es eliminar el *sesgo de selección* o la distorsión de los datos que surge de la forma como se recopilan. Si no se tiene en cuenta, puede resultar un sesgo de selección en cualquier conclusión que se saque por estar equivocados los datos. Como quiera que todas las personas o las cosas tienen igual probabilidad de ser seleccionadas, se elimina el sesgo de selección. Existen diferentes formas de sesgo de selección, una de ellas que es común es el sesgo de autoselección, que a su vez tiene también varias formas. El sesgo de autoselección ocurre en cualquier programa para el cual las personas se presentan de manera voluntaria, e incluso se encuentra en las personas que responden voluntariamente a una encuesta. La fuente de sesgo es que los voluntarios y encuestados a menudo comparten características que no se encuentran en aquellos que no se presentan como voluntarios, o que no responden la encuesta.

Muestra aleatoria: Muestra en la que cada unidad de la población tiene igual probabilidad de ser seleccionada.

Sesgo de selección: Distorsión de la evidencia causada por la forma como se recopilan los datos.

Una muestra aleatoria debe ser representativa de la población como un todo. Dicha muestra permite a los evaluadores hacer generalizaciones sobre la población de la cual se extrajo.

Marco de muestra: Conjunto completo de unidades del cual se extrae una muestra.

Una lista completa de cada unidad de la población de interés, llamada un *marco de muestra*, es necesaria para seleccionar una muestra aleatoria. A cada unidad de la población se le asigna un número único, y luego se seleccionan las unidades de toda la población, utilizando un programa aleatorio, por ejemplo, una tabla de números aleatorios que contiene conjuntos de dígitos (por ejemplo, 0, 1, 2, 3, …) organizados en orden aleatorio. Las fuentes de tablas de números aleatorios en la web se suministran en las referencias al final de este capítulo. Muchas personas utilizan un generador basado en la web para producir números aleatorios. Las fuentes como Stattrek preguntan cuántos números aleatorios se necesitan, y cuáles son los valores mínimo y máximo, y luego generan los números aleatorios.

Hay seis tipos de muestras aleatorias:

- Aleatorias simples.
- De aleatoria de intervalos.
- Con arranque aleatorio y seguido de intervalos fijos.
- Aleatorias estratificadas.
- De grupos aleatorios.
- Aleatorias de etapas múltiples.

Muestras aleatorias simples

Una muestra aleatoria simple es el tipo más común y más sencillo de muestra aleatoria. Se utiliza cuando el objetivo primario es hacer inferencias sobre la población total y no sobre subpoblaciones específicas.

Las muestras aleatorias simples son buenas para extraer muestras de entre 50 y 500 unidades de poblaciones homogéneas, o muestras mayores de poblaciones más heterogéneas (Jackson, 2008). Se pueden extraer de varias formas. Por ejemplo, los evaluadores interesados en observar actividades en el salón de clases para medir la cantidad de tiempo que los estudiantes emplean en actividades de práctica de aprendizaje, podrían seleccionar al azar los salones, las horas del día o los días de la semana. Los evaluadores que desean observar el volumen de tráfico en la vía de la aldea a un pueblo mayor, podrían seleccionar al azar las horas del día o los días de la semana, las veces por año, o los puntos de observación a lo largo de la vía.

El procedimiento para extraer una muestra aleatoria simple es como sigue (Jackson, 2008):

1. Definir cuidadosamente la población, indicando lo que incluye y lo que excluye.
2. Encontrar o generar un marco de muestra que relacione todas las unidades de la población, y asignar un número único a cada una. Los números no tienen que ser contiguos (consecutivos), pero será más rápida la extracción de la muestra sin espacios grandes entre los números asignados.
3. Decidir el tamaño de la muestra.
4. Determinar el número de dígitos del número mayor asignado.

5. Adquirir una tabla de números aleatorios.
6. Decidir el patrón fijado para leer los números (por ejemplo, de arriba abajo, y luego la columna siguiente, leyendo los primeros dígitos).
7. Seleccionar a ciegas un punto de partida en la tabla.
8. Leer los números en el orden especificado por el patrón seleccionado, leyendo siempre el número de dígitos establecido en el numeral 4.
9. Cada vez que se lea un dígito que corresponda con el número de una unidad del marco de muestra, marcar esa unidad como seleccionada.
10. Continuar hasta haber seleccionado el número de unidades que se requiere.
11. Si se obtiene un número aleatorio igual a otro ya utilizado, escoger el número siguiente. No puede seleccionarse la misma unidad dos veces.

O, por ejemplo, un equipo de evaluación selecciona 100 archivos de una población de 500. Todos los archivos se numeraron de manera consecutiva de 001 a 500, y se registraron en orden numérico. El equipo puede utilizar entonces un generador de números aleatorios, indicando que se requieren 100 números, que el valor mínimo es 001 y el máximo 500.

Muestreo de intervalos aleatorios y muestreo con arranque aleatorio y seguido de intervalos fijos

El *muestreo aleatorio de intervalos* puede utilizarse cuando hay una población secuencial que no está enumerada, y sería difícil enumerarla o tomaría mucho tiempo hacerlo. A semejanza de una muestra aleatoria simple, una de intervalos aleatorios utiliza una tabla de números aleatorios. Sin embargo, en lugar de utilizar la tabla para seleccionar las unidades de la muestra, se utiliza para seleccionar el intervalo (la cuenta entre los números) sobre el cual seleccionar las unidades.

Muestreo de intervalo aleatorio: Muestra seleccionada utilizando distancias entre números escogidas al azar.

Para extraer una muestra aleatoria de intervalos se utilizan los pasos siguientes (Jackson, 2008):

1. Estimar el número de unidades de la población.
2. Determinar el tamaño deseado de la muestra.
3. Dividir el número estimado de unidades entre el tamaño deseado de la muestra. Redondear este resultado a dos decimales para obtener la longitud promedio del intervalo aleatorio, requerida para producir el tamaño deseado de la muestra al tomar la muestra de la población completa.
4. Multiplicar el resultado obtenido en el paso 3 por un quinto y redondear a dos decimales para obtener el "multiplicador".
5. Designar a ciegas un punto de partida en la población.
6. Utilizar una tabla de números aleatorios para seleccionar un número aleatorio de un dígito, multiplicarlo por el multiplicador, redondear al entero más próximo, contar desde el punto de partida hacia abajo ese número de lugares, seleccionar esa unidad para la muestra y colocar un marcador donde fue seleccionado.

7. Tomar el siguiente número aleatorio de un dígito, multiplicarlo por el multiplicador, redondearlo al entero más próximo y contar esos lugares más.
8. Continuar en la misma forma a través de toda la población hasta llegar al punto de inicio.
9. *Advertencia:* Si es importante devolver las unidades a sus posiciones originales, asegurarse de marcar cada lugar del que se seleccionó un archivo con un identificador, de manera que pueda devolver los archivos a sus lugares apropiados.

Muestreo con arranque aleatorio y seguido de intervalos fijos: Técnica sistemática de muestreo en la que el punto de partida es aleatorio, pero luego se selecciona cada *n*-ésimo caso.

Los muestreos de intervalos aleatorios pueden generarse también utilizando programas de computador.

Algunas veces no es posible realizar una selección verdaderamente aleatoria. En este caso, puede utilizarse una técnica de muestreo sistemática, empezando desde un punto aleatorio y luego seleccionar cada *n*-ésimo caso.

En el muestreo con arranque aleatorio y seguido de intervalos fijos se consideran algunas veces muestras cuasialeatorias, porque el punto de partida es aleatorio pero el intervalo no lo es. Aunque esta técnica de muestreo es algo más simple que el muestreo de intervalos aleatorios, adolece de una seria debilidad en condiciones particulares. Por ejemplo, considérese una evaluación en la que se requiere una muestra diaria de registros de un mercado que abre todos los días de la semana. La población se estima en 700 registros y la muestra debe ser de 100. En este caso, el intervalo fijo será cada séptimo registro, lo que daría como resultado que todos los registros se seleccionen para el mismo día de la semana, lo cual no es probable que sea representativo de todos los días de la semana. Para evitar esa situación, si existe alguna posibilidad de que un intervalo fijo pueda producir una muestra sesgada, debe utilizarse el procedimiento de muestreo de intervalos aleatorios.

Un muestreo con arranque aleatorio y seguido de intervalos fijos se selecciona en la forma siguiente:

1. Estimar el número de unidades de la población.
2. Determinar el tamaño deseado de la muestra.
3. Dividir el resultado del paso 1 entre el del paso 2.
4. Designar a ciegas un punto de partida en la población.
5. Contar hacia abajo el intervalo especificado y seleccionar esa unidad para la muestra.
6. Continuar contando hacia abajo el mismo intervalo y seleccionar las unidades hasta que se regrese al punto de partida.

Muestras aleatorias estratificadas

Con frecuencia, es importante incluir grupos específicos en una muestra que puede estar representada en forma deficiente, utilizando una muestra aleatoria simple. Para

Muestra aleatoria estratificada: Muestra en la que el marco de muestra se divide en dos o más estratos (subpoblaciones), de los cuales se seleccionan los participantes al azar.

asegurar el muestreo de tales grupos, los evaluadores dividen la población en estratos basados en alguna característica significativa, como género, edad, grupo étnico u otra característica, y seleccionar la muestra. Esta clase de muestra se denomina *muestra aleatoria estratificada.*

Supóngase que los evaluadores desean analizar los efectos de una intervención sobre residentes urbanos y rurales. Si los rurales representan solamente una pequeña proporción de la población total en la zona, una muestra aleatoria simple no puede incluir los suficientes para realizar un análisis significativo.

Una muestra aleatoria estratificada se toma dividiendo la población en grupos no superpuestos (estratos) (n_1, n_2, n_3, …, n_i, tales que $n_1 + n_2 + n_3 + \ldots + n_i = n$), y tomando una muestra aleatoria dentro de cada estrato. El número seleccionado de cada estrato debe ser equivalente a la proporción del estrato del total de la población.

Muestra aleatoria de conglomerados

Muestra aleatoria de conglomerados: Muestra tomada de grupos que ocurren naturalmente de la unidad de análisis.

Una *muestra aleatoria de conglomerados* es una muestra tomada de grupos que ocurren naturalmente de la unidad de análisis. Las familias (o los hogares) son grupos de personas, los pueblos son grupos de familias. Los grupos deben ser mutuamente excluyentes y colectivamente exhaustivos. Las muestras de grupos se utilizan mucho en las situaciones siguientes:

- No hay una lista completa de toda la población de interés, pero hay una lista completa de los grupos en que aquéllas ocurren.
- Hay una lista completa, pero en ella los nombres están tan dispersos que tomaría demasiado tiempo y muchos costos enviar a los recopiladores de datos para efectuar una muestra aleatoria simple.

En una muestra aleatoria de conglomerados, se hace el muestreo del grupo de manera aleatoria, y se recopilan los datos sobre todas las unidades objetivo. Esta técnica es particularmente útil cuando las unidades objetivo se encuentran localizadas a grandes distancias entre sí (recuadro 9.1).

El principal inconveniente de las muestras aleatorias de conglomerados es que probablemente produzcan estimados menos precisos del parámetro de la población (característica) que las muestras aleatorias simples o las estratificadas del mismo número de unidades de interés. Puede que las clínicas seleccionadas, mencionadas en el recuadro 9.1, atiendan a pacientes cuyas características económicas o religiosas sean diferentes de los que no queden incluidos en la muestra. De ser así, los resultados de la muestra producirán estimados sesgados de la población total de pacientes de Sida, atendidos por las clínicas de la región.

Recuadro 9.1 *Usar muestreo de grupos para identificar personas con Sida y entrevistarlas*

Un equipo de evaluación busca entrevistar 200 personas con Sida que viven en una región. No existe una lista de personas con Sida en la región, y sería prohibitivamente costoso intentar recopilar dicha lista, y además el proceso de obtenerla podría suscitar algunos problemas éticos y de confidencialidad. La región cuenta, sin embargo, con 25 clínicas de salud y los evaluadores saben que la mayoría de ellas atiende a unos 50 pacientes de Sida, por lo que decidieron efectuar un muestreo aleatorio de las clínicas y estudiar todos los pacientes de Sida de las clínicas de la muestra. Las 25 clínicas de salud están repartidas en una gran región cuyas vías son deficientes. Los evaluadores saben que la mayoría de las clínicas atienden a unos 50 pacientes de Sida y, en consecuencia, hacen una muestra aleatoria de 4 de las 25 clínicas, y estudian todos los pacientes de Sida de esas clínicas, con lo que el tamaño de la muestra es de unos 200 pacientes.

Muestras aleatorias de etapas múltiples

Muestra aleatoria de etapas múltiples: Muestra en la que se utilizan en forma secuencial dos o más procedimientos de muestreo aleatorio.

En la muestra aleatoria de etapas múltiples se combinan dos o más formas de muestreo aleatorio. El proceso empieza usualmente con muestreo de grupos aleatorio, y luego se aplica el muestreo aleatorio simple o el estratificado. En el ejemplo del recuadro 9.1, con una muestra aleatoria de etapas múltiples se tomaría inicialmente una muestra de grupos de 8 clínicas, y luego se tomaría una muestra aleatoria simple de 25 pacientes de cada clínica. Con esta técnica se obtendría una muestra de 200 pacientes, como en el ejemplo anterior, pero provendrían de un número mayor de clínicas.

Se puede combinar el muestro aleatorio con el muestreo no aleatorio en una muestra de etapas múltiples. Por ejemplo, puede efectuarse el muestreo de las clínicas del recuadro 9.1 en forma no aleatoria, y después tomar una muestra de los pacientes de Sida en forma aleatoria en cada clínica.

El inconveniente de las muestras de etapas múltiples y de grupos es que proporcionan estimados algo menos precisos de la población que las muestras aleatorias simples. En el ejemplo de las clínicas de Sida, solamente se toman muestras aleatorias de 8 clínicas del total de 25, y pueden no ser representativas de las 25.

Muestreo no aleatorio

Cuando no es posible efectuar el muestreo aleatorio, debe utilizarse un enfoque distinto. Las muestras no aleatorias se clasifican usualmente en tres tipos: bien dirigidas, de bola de nieve y de conveniencia.

Las técnicas de muestreo no aleatorio solamente permiten a los evaluadores una capacidad limitada para generalizar los hallazgos a la población general. Dichas muestras pueden, sin embargo, robustecer los resultados de una evaluación.

Muestras bien dirigidas

En una *muestra bien dirigida* (llamada también *muestra de juicio*), las selecciones se efectúan de modo que cumplan fines específicos de un estudio. La selección se basa en criterios predeterminados que, a juicio de los evaluadores, proporcionarán los datos necesarios.

Las siguientes son las formas más utilizadas de muestras bien dirigidas (Jackson, 2008):

- ***Muestra de caso típico (mediana)***: se toman deliberadamente las unidades de las características usuales o más comunes de una población (las que se encuentran en el rango medio de una curva de campana). El propósito del estudio es observar muy de cerca los elementos usuales y no los atípicos.
- ***Muestra máxima de variación (heterogeneidad)***: se toman las unidades de modo que representen el rango completo de una característica de interés. Se obtienen de todas partes de la curva de campana.
- ***Muestra de cuota***: se toman las unidades de modo que haya igual número o igual proporción de cada estrato. Por ejemplo, el evaluador elige cinco unidades del tercio superior, cinco del tercio de la mitad y cinco del tercio inferior de una distribución.
- ***Muestra de caso extremo***: se toman las unidades de los casos extremos de una distribución. Se seleccionarían de los extremos izquierdo y derecho de una curva de campana. Las muestras de caso extremo corresponden a las unidades menos comunes.
- ***Muestra de desconfirmación de casos:*** se seleccionan las unidades de casos que se sabe confirman o contradicen la sabiduría, los principios o la teoría convencional (por ejemplo, de proyectos bien preparados y exitosos y de proyectos bien preparados y fallidos).

Muestra bien dirigida: Muestra en la cual las selecciones se hacen con base en criterios preestablecidos.

Muestra de caso típico: Muestra bien dirigida en la cual se toman las unidades del rango medio de la distribución.

Muestra de variación máxima: Muestra bien dirigida en la cual se toman las unidades para representar un rango completo de un parámetro.

Muestra de cuota: Muestra bien dirigida en la cual se selecciona un número específico de distintos tipos de unidades.

Muestra de caso extremo: Muestra bien dirigida en la cual se toman las unidades de los extremos izquierdo y derecho de una distribución.

Muestreo en bola de nieve

Los muestreos en bola de nieve (conocidas también como *muestras de referencia en cadena*) se utilizan cuando los evaluadores no saben a quién o qué incluir en una muestra y cuando se desconocen las demarcaciones de la población y no hay un marco de muestra.

El muestreo en bola de nieve se utilizan a menudo en las entrevistas. Los evaluadores que utilizan esta clase de muestra bien dirigida preguntan quiénes satisfacen los criterios de inclusión en el estudio con quién más deberían hablar.

Muestra de desconfirmación de casos: Muestra bien dirigida en la cual las unidades se toman de casos conocidos para confirmar o contradecir la sabiduría, los principios o la teoría convencional.

El muestreo en bola de nieve contienen serios sesgos potenciales al depender de una cadena de referencias. En Nambiar (2008) se describe un ejemplo de su uso. Deben utilizarse con cautela.

Muestras de conveniencia

En una *muestra de conveniencia*, se hacen las selecciones a conveniencia del evaluador. Los ejemplos comunes de muestras de conveniencia en la evaluación para el desarrollo incluyen:

Muestras de conveniencia: Cuando los evaluadores no saben a quién o qué incluir en una muestra se hacen las selecciones a conveniencia del evaluador.

- Visitar los sitios del proyecto más cercanos a un aeropuerto.
- Entrevistar a los administradores del proyecto que se encuentren disponibles el día de la visita del evaluador.
- Observar las zonas físicas que los funcionarios del proyecto elijan mostrar.
- Hablar con los representantes de una ONG y con los líderes del pueblo con los que se encuentre el evaluador (Jackson, 2008).

Las muestras de conveniencia constituyen un medio extremadamente débil de inferir cualquier clase de evidencia o de patrón de los datos, pues no hay forma de saber qué tan diferentes son estas muestras de la población pertinente como un todo.

Deficiencias del muestreo no aleatorio

Aunque las muestras no aleatorias tienen usos legítimos en la evaluación para el desarrollo, tienen también varias deficiencias:

- Algunas veces los autores de las políticas las perciben como menos confiables que las muestras aleatorias.
- Dado que no satisfacen los supuestos de la estadística inferencial, nunca deberían aplicarse a muestras no aleatorias pruebas de significación estadística y cómputos de intervalos de confianza.
- Las muestras no aleatorias están sujetas a toda clase de sesgos.

Sin muestreo aleatorio no pueden hacerse generalizaciones para una población mayor. Sin embargo, los datos pueden ser muy útiles y lo mejor disponible dadas las restricciones.

Muestreo combinado

Los métodos de muestreo aleatorio y no aleatorio pueden combinarse. Por ejemplo, un equipo de evaluación que recopila datos sobre colegios, puede seleccionar dos colegios de las comunidades más pobres y dos de las más ricas. De estos cuatro colegios podrían entonces seleccionar los estudiantes para su muestra.

Establecer el tamaño de la muestra

Aun cuando se utilice una muestra aleatoria, existe alguna posibilidad de error, pues la muestra puede ser diferente de la población. Las estadísticas se utilizan para estimar la probabilidad de que los resultados de la muestra sean representativos de la población total. Los estadísticos han desarrollado teorías y fórmulas para hacer estos estimados y determinar el tamaño apropiado de la muestra. Esta sección se centra en el conocimiento de los conceptos básicos del análisis estadístico, y cómo aplicarlos al diseño de evaluaciones.

Al escoger el *tamaño de la muestra,* los evaluadores deben decidir qué nivel de confianza requieren para que los resultados de la muestra reflejen con precisión la población total pertinente. El nivel de confianza utilizado generalmente es el 95%, el cual significa que 95 veces de cada 100 los resultados de la muestra reflejarán con precisión a la población total. Si los evaluadores están dispuestos a tener una certeza del 90%, podrían utilizar una muestra menor. Si desean una certeza del 99% necesitarían una muestra mayor.

Tamaño de la muestra: Número de observaciones de una muestra.

Una vez escogen el tamaño de la muestra, los evaluadores deben a continuación determinar la precisión de sus estimados. A diferencia del sesgo, el error muestral puede calcularse y es la diferencia promedio entre todos los estimados para todas las muestras posibles, y el valor obtenido de haberse estudiado la población total. Esto se denomina algunas veces *error de muestreo* o margen de error. Una encuesta puede informar que el 48% de la población está a favor de que se eleven los impuestos y el 52% en contra, con un error muestral de +/-3 puntos porcentuales. Lo que esto significa es que si toda la población fuese encuestada, las proporciones reales serían de entre 45% y 51% (48 +/-3) a favor y entre 49% y 55% (52 +/-3) en contra de elevar los impuestos. Los ±3 puntos porcentuales son lo que se denomina un *intervalo de confianza* (que no debe confundirse con el nivel de confianza). Nótese que los dos intervalos se traslapan parcialmente, lo cual significa que no se puede estar seguro, con un nivel de confianza del 95%, de que más personas en la población se oponen a la subida de los impuestos que las que están a favor de ella. Los resultados de la muestra son demasiado cerrados para decirlo.

Error de muestreo (o margen de error): Estimado de error causado por observar una muestra y no el total.

Intervalo de confianza: Rango dentro del cual cae el verdadero valor de la población con una probabilidad dada.

El tamaño de la muestra es función del tamaño de la población de interés y los niveles de confianza y de precisión

deseados. El tamaño apropiado de la muestra puede establecerse de dos maneras: la primera es utilizar una fórmula, y la segunda es utilizar una tabla que indique el tamaño de muestra requerido para un nivel de confianza dado (cuadro 9.1).

Cuanto más pequeña sea la población, mayor será el tamaño de la muestra relativo a la población como un todo. Por ejemplo, si la población es 300, se necesita un tamaño de muestra de 169 –apenas por encima de la mitad de la población total– para obtener un nivel de confianza del 95%. Una población de 900 requeriría un tamaño de muestra de 269, o menos de un tercio de la población. Si la población excede 100.000, se necesitaría una muestra de 385, apenas un 0,385% de la población total.

Al tomar muestras de unidades (por ejemplo, distritos, colegios, maestros, padres o ciudadanos) para participación voluntaria en una evaluación, no hay seguridad de encontrar a todos los seleccionados (aun con el marco de muestra más actualizado), o de que respondan a la solicitud de participación, o de que respondan afirmativamente para participar en la evaluación. Por este fenómeno, muchos evaluadores escogen una muestra mayor de la necesaria para obtener el tamaño de la muestra que necesitan. Cuando la tasa de respuesta real difiere de la planeada en un 10% o más, la probabilidad del *sesgo de respuesta* es alta y requiere análisis (*ver* el capítulo 8, "Seleccionar y construir instrumentos de recopilación de datos").

Cuadro 9.1 *Tamaño mínimo requerido de la muestra al nivel de confianza del 95% con un margen de error de 5 puntos porcentuales*

Tamaño de la población	*Tamaño de la muestra*	*Tamaño de la población*	*Tamaño de la muestra*
10	10	550	226
20	19	600	234
40	36	700	248
50	44	800	260
75	63	900	269
100	80	1.000	278
150	108	1.200	291
200	132	1.300	297
250	152	1.500	306
300	169	3.000	341
350	184	6.000	361
400	196	9.000	368
450	207	50.000	381
500	217	100.000+	385

Fuente: Krejcie y Morgan, 1970.

Tasa de respuesta: Porcentaje de la muestra deseada de la que se recopilan realmente los datos.

En el cuadro 9.2 se presentan los tamaños de muestra para poblaciones muy grandes (1 millón y más). En muchas encuestas nacionales se utilizan muestras de alrededor de 1.100, lo que deja un margen de error de +/-3 puntos porcentuales con un nivel de confianza del 95%.

Las herramientas disponibles en Internet identifican el tamaño de población requerido para lograr un nivel de confianza y un margen de error dados. Los vínculos con estas fuentes se encuentran al final del presente capítulo.

Para resumir:

- Pueden mejorarse la exactitud y precisión, aumentando el tamaño de la muestra. Para aumentar la exactitud y reducir el margen de error, aumentar el tamaño de la muestra.
- El estándar para lo que se debe lograr es un nivel de confianza del 95% (margen de error de +/-5%).
- A mayor margen de error, menor precisión en los resultados.
- Cuanto más pequeña la población, mayor la razón del tamaño de la muestra al tamaño de población.

Si una evaluación requiere una compleja estrategia de muestreo de etapas múltiples, los evaluadores posiblemente solicitarán ayuda. La American Statistical Association tiene un directorio de consultores estadísticos, y la Alliance of Statistics Consultants ofrece asistencia para el manejo y el análisis de datos, la consulta de tesis y capacitación y tutoría en estadística. HyperStat Online proporciona vínculos para muchos otros recursos que ofrecen asistencia en estadística. Los vínculos con estas páginas web se encuentran al final de este capítulo.

Cuadro 9.2 *Tamaños de muestra para poblaciones de 1 millón y más*

Margen de error (porcentaje)	*Nivel de confianza*		
	99%	*95%*	*90%*
± 1	16.576	9.604	6.765
± 2	4.144	2.401	1.691
± 3	1.848	1.067	752
± 5	666	384	271

Fuente: Jackson, 2007.

Resumen

Usualmente es imposible o poco práctico recopilar datos de cada fuente de datos. Por esto los evaluadores recopilan solamente una muestra de las fuentes.

Hay dos formas de muestreo, el aleatorio y el no aleatorio. Las muestras aleatorias son aquellas en las cuales cada unidad de la población tiene igual probabilidad de ser seleccionada. En el muestreo aleatorio se examina un subconjunto de la población, y luego se generalizan los resultados a la población total.

Existen seis tipos de muestras aleatorias:

- Aleatorias simples.
- Intervalos aleatorios.
- Con arranque aleatorio y seguido de intervalos fijos.
- Aleatorias estratificadas.
- Aleatoria de conglomerados.
- Aleatorias de etapas múltiples.

Existen tres clases de muestras no aleatorias:

- Bien dirigidas.
- De bola de nieve.
- De conveniencia.

No es necesario utilizar solamente una clase de muestreo en las evaluaciones, y pueden utilizarse combinaciones de técnicas de muestreo.

El tamaño de la muestra es función del tamaño de la población de interés, y los niveles de confianza y de precisión deseados. Aun cuando se utilice una muestra aleatoria, existe alguna posibilidad de error. Los evaluadores utilizan varios métodos estadísticos para fijar el nivel y el intervalo de confianza.

Actividades del capítulo 9

Ejercicio de aplicación 9.1: muestreo

El pueblo que usted está estudiando tiene una población de 300 habitantes, todos los cuales están registrados en el pueblo, cada uno con un número único de identificación. Cuenta con el tiempo y los recursos para tomar una muestra de hasta 50 habitantes. Usted desea hacer generalizaciones de la población total de 100 a partir de la muestra. ¿Es posible esto? ¿Qué tipo de muestra elegiría? Justifique su decisión.

44	14	12	12	03	12	73	72	62	33	35	62	80	34	77
69	59	54	90	01	50	04	93	76	69	43	95	47	60	80
23	95	24	95	24	55	69	89	41	18	12	94	43	21	43
40	76	50	38	18	05	44	23	72	61	58	67	99	05	75
54	05	51	52	04	34	25	64	90	95	02	86	51	14	37
36	82	03	65	38	93	49	64	06	93	01	30	62	05	68
96	19	97	24	16	26	94	14	17	45	22	51	09	92	16
75	85	18	50	50	60	80	52	42	11	05	70	89	53	38
57	78	12	98	55	51	48	77	54	07	66	15	33	44	64
58	20	10	51	62	06	25	56	63	67	73	73	79	05	65
55	84	17	67	52	38	16	29	05	24	12	05	35	87	31
92	44	84	04	17	47	18	78	54	40	02	59	74	06	73
86	96	79	86	75	67	31	41	40	20	87	17	85	98	70
78	84	03	69	43	38	43	98	90	75	56	49	88	52	78
25	05	76	72	06	59	37	56	24	36	95	05	30	62	02
26	67	04	13	77	37	21	57	77	41	82	30	32	80	09

Ejercicio de aplicación 9.2: uso de la tabla de números aleatorios

La pequeña aldea que está estudiando tiene 90 familias. Usted tiene una lista numerada de ellas, de la cual desea seleccionar una muestra aleatoria simple de 10. ¿Cómo puede utilizar la tabla de números aleatorios siguiente para seleccionar la muestra?

Ejercicio de aplicación 9.3: estrategia de muestreo

Trabajando en pequeños grupos, si es posible, identificar una medida o estadística apropiada para cada una de las preguntas de evaluación siguientes. Luego decidir qué estrategia simple escogería para cada una de estas situaciones y explique por qué la escogería.

1. ¿Cuál es la calidad de las vías en las aldeas al noroeste de Camboya, inmediatamente después de la estación lluviosa?
2. ¿Qué proporción de niños en Tamil Nadu contrae malaria al menos una vez antes de cumplir 10 años?
3. ¿Cuáles son las características demográficas de las personas que visitan clínicas de salud rurales en Sri Lanka?

Referencias y lecturas adicionales

Easton, V. J. y J. H. McColl. 2007. *Statistics Glossary: Sampling.* http://www.stats.gla.ac.uk/steps/glossary/sampling.html.

Guba, E. y Y. S. Lincoln. 1989. *Fourth Generation Evaluation.* Thousand Oaks, CA: Sage Publications.

Henry, G. T. 1990. *Practical Sampling.* Thousand Oaks, CA: Sage Publications.

Jackson, Gregg B. 2007. *Sampling for IEG Managers.* Presentación en George Washington University, Washington, DC, diciembre 18.

———. 2008. *Sampling in Development Evaluations.* International Program for Development Evaluation Training (IPDET); presentación, Ottawa, junio 30 y julio 1.

Kish, L. 1995. *Survey Sampling.* Nueva York: John Wiley and Sons.

Krejcie, R. V. y D. W. Morgan. 1970. "Determining Sample Size for Research Activities". *Educational and Psychological Measurement* 30: 607-10.

Kumar, R. 1999. *Research Methodology: A Step-by-Step Guide for Beginners.* Thousand Oaks, CA: Sage Publications.

Laws, S., with C. Harper y R. Marcus 2003. *Research for Development: A Practical Guide.* Thousand Oaks, CA: Sage Publications.

Levy, P. y S. Lemeshaw. 1999. *Sampling of Populations.* 3ª ed. Nueva York: John Wiley and Sons.

Lipsey, M. W. 1990. *Design Sensitivity: Statistical Power for Experimental Research.* Thousand Oaks, CA: Sage Publications.

Lohr, S. 1998. *Sampling: Design and Analysis.* Pacific Grove, CA: Duxbury Press.

Nambiar, Devaki. 2008. "The Delhi Snowball: Sampling Escapades in Urban India". Documento presentado en la Annual Meeting of the International Communication Association, Montreal, mayo 22-26.

Neuman, W. Lawrence. 2006. *Social Research Methods: Qualitative and Quantitative Approaches.* 6ª ed. Boston: Allyn and Bacon.

Patton, M. Q. 2002. *Qualitative Research and Evaluation Methods.* Thousand Oaks, CA: Sage Publications.

Scheyvens, R. y D. Storey, eds. 2003. *Development Fieldwork: A Practical Guide.* Thousand Oaks, CA: Sage Publications.

Stattrek. http://stattrek.com/tables/randon.asph.

Tryfos, P. 1996. *Sampling Methods for Applied Research.* Nueva York: John Wiley and Sons.

Sitios web

Alliance of Statistics Consultants. http://www.statisticstutors.com/#statisticalanalysis.

American Statistical Association Directory. http://www.amstat.org/consultant directory/index.cfm.

Dr. Drott's Random Sampler. http://drott.cis.drexel.edu/sample/content.html.

HyperStat Online. "Chapter 11: Power". http://davidmlane.com/hyperstat/power.html.

HyperStat Online: Help with Statistics: Statistical Consultants and Tutors. http://davidmlane.com/hyperstat/consultants.html.

Probability Sampling. http://www.socialresearchmethods.net/kb/sampprob.htm.

Power Analysis. http://www.statsoft.com/textbook/stpowan.html.

Research Randomizer. http://www.randomizer.org.

StatPages.net. Web Pages that Perform Statistical Calculations, http://www.StatPages.net.

Survey Research Methods Section. http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html.

The Survey System: Sample Size Calculator. http://www.surveysystem.com/sscalc.htm.

UCLA Statistics Calculator: http://calculators.stat.ucla.edu.

001El camino.indd 350 08/11/2010 03:27:55 p.m.

CAPÍTULO 10

Planear y hacer el análisis de datos

Una vez recopilados los datos, los evaluadores deben examinarlos para encontrar su significado. Este proceso empieza con una estrategia de análisis de datos. Los datos cualitativos y cuantitativos demandan diferentes estrategias y técnicas.

Estrategia de análisis de datos

Desarrollar la estrategia de análisis de datos es una parte importante del proceso de planeación. Los evaluadores deben estar conscientes de las opciones de análisis de datos –y de sus respectivas fortalezas y debilidades– al planear la evaluación. En la matriz de diseño, el objetivo debe ser específico e indicar el análisis y los gráficos que resulten de la información recopilada. Un error común es recopilar grandes cantidades de datos que nunca se utilizan.

Ya sea que el diseño de la evaluación haga énfasis principalmente en datos cualitativos o en datos cuantitativos, la recopilación y el análisis de los datos se traslaparán. Al principio de la recopilación de datos se emplea un corto tiempo en el análisis de los datos, especialmente si primero se va a efectuar una prueba piloto. Al continuar la evaluación se emplea mayor tiempo en el análisis de los datos y menos en su recopilación.

001El camino.indd 351 08/11/2010 03:27:55 p.m.

El análisis cualitativo es apropiado, por ejemplo, en situaciones en las cuales se utiliza una guía de entrevista semiestructurada para obtener una percepción profunda de una intervención. Se utilizaría para analizar respuestas a preguntas como estas:

- ¿Qué dificultades encuentra el personal?
- ¿Por qué dicen los participantes que abandonaron pronto?
- ¿Cómo es la experiencia para los participantes?

El análisis cuantitativo debe utilizarse para responder preguntas para las cuales se utilizaron instrumentos de recopilación de datos estructurados, por ejemplo, una encuesta. Podría utilizarse para responder preguntas como las siguientes:

- ¿Cuáles son los puntajes promedio para los distintos grupos de participantes?
- ¿Cómo califican los participantes la aplicabilidad de la intervención en una escala de uno a cinco?
- ¿Cuánta variabilidad hay en las respuestas a la pregunta?
- ¿Son estadísticamente significativas las diferencias entre los dos grupos?

Análisis de datos cualitativos

Análisis de datos cualitativos: Procedimiento utilizado para analizar información recopilada en forma no numérica, por ejemplo, una narrativa escrita o respuestas grabadas en entrevistas semiestructuradas y observaciones u otros documentos y medios, para comprender e interpretar comportamiento y situaciones; el procedimiento incluye la identificación de los temas.

Se utiliza para explicar los datos no numéricos recopilados como parte de la evaluación. El análisis de observaciones semiestructuradas, entrevistas abiertas, documentos escritos y transcripciones de grupos de discusión requiere usar técnicas cualitativas.

El análisis de datos cualitativos empieza desde que se está en el campo, cuando pueden surgir percepciones. Parte del trabajo de campo es grabar y registrar percepciones analíticas que ocurren durante la recopilación de datos, la cual debería traslaparse con el análisis de datos, mientras el evaluador tenga cuidado de no permitir interpretaciones iniciales para limitar las posibilidades analíticas (cuadro 10.1). Según anota Patton (2002, p. 436):

> Para la recopilación de datos basados en encuestas, pruebas estandarizadas y diseños experimentales, la línea entre la recopilación y el análisis de los datos es clara. Pero la naturaleza fluida y emergente de la investigación naturalista hace que la distinción entre la recopilación y el análisis de los datos sea mucho menos absoluta. Durante el trabajo de campo surgen ideas sobre direcciones para el análisis, los patrones toman forma y posibles temas brotan en la mente. Surgen hipótesis que aportan información al trabajo de campo posterior. Mientras las primeras etapas del trabajo de campo suelen ser generativas y emergentes, siguiendo a dondequiera que lleven los datos, en las etapas posteriores se finaliza esta labor y se pasa a la recopilación de datos confirmatorios, profundizando las percepciones de los patrones que hayan aparecido y confirmándolos (o contradiciéndolos).

A veces se encuentran vacíos o ambigüedades en el análisis. Si la programación, el presupuesto y otros recursos lo permiten, un evaluador puede regresar al campo a recopilar más datos, a fin de clarificar respuestas o hacer nuevas observaciones. Al recopilar los datos, los integrantes de un equipo dialogan diaria o semanalmente para discutir nuevos temas y adaptar protocolos, si es indicado.

Tomar buenas notas

Al recopilar datos cualitativos, es importante capturar con precisión todas las observaciones; para eso son esenciales las buenas notas. Esto significa que debe prestarse mucha atención a lo que la gente dice y cómo lo dice. Al tomar notas, los evaluadores deben tratar de no interpretar lo que la gente dice, sino escribir lo que observan, incluyendo el lenguaje corporal y cualquier cosa potencialmente pertinente que suceda durante la recopilación de datos (por ejemplo, interrupciones durante una entrevista). Los evaluadores deberían captar pensamientos, reacciones e interpretaciones inmediatas y mantenerlas en sección aparte en sus notas.

Cuadro 10.1 *Consejos para recopilar y analizar datos cualitativos*

Tarea	*Consejos*
Recopilar datos	• Mantener buenos registros
	• Escribir las entrevistas, impresiones y notas de los grupos de discusión inmediatamente después de recopilarlos.
	• Hacer comparaciones constantes a medida que se avanza.
	• Reunirse regularmente con el equipo para comparar notas, identificar temas y hacer ajustes.
Resumir datos	• Escribir resúmenes de una página inmediatamente después de cada entrevista o grupo de discusión.
	• Incluir todos los temas principales.
	• Identificar el asunto o la información obtenida más interesantes, ilustrativos o importantes.
	• Identificar nuevas preguntas para explorar.
Utilizar herramientas para seguimiento	• Crear un archivo separado para las propias reacciones durante el estudio, incluidos los sentimientos, corazonadas y reacciones.
	• Anotar las ideas propias a medida que surgen.
	• Mantener un archivo de citas del proceso de recopilación de datos para su uso al animar la narrativa cuando se escriba el informe.
Almacenar datos	• Asegurarse de que toda la información se encuentre en un solo lugar.
	• Hacer copias de toda la información y colocar los originales en un archivo central.
	• Utilizar las copias para escribir, cortar y pegar según sea necesario.

Fuente: los autores.

Triangulación: Uso de tres o más teorías, fuentes, tipos de información o de análisis para verificar y sustanciar una evaluación, haciendo verificaciones cruzadas de los resultados.

Es importante dar tiempo a los evaluadores inmediatamente después de una entrevista, una observación, un ejercicio de cartografía o un grupo de discusión, para revisar, escribir y hacer adiciones a sus notas, de modo que puedan encontrar explicaciones sobre ellas más adelante. Es sorprendente lo difícil que se hace comprender las notas tomadas incluso sólo un día antes si no se escriben con claridad.

Aun si se graba la sesión, los evaluadores deben invertir por lo menos algún tiempo en una escritura preliminar mientras la sesión está todavía fresca. Esto puede ahorrar horas de escucha u observación de cintas, o sumergirse en transcripciones.

La *triangulación* es el uso de tres o más teorías, fuentes, tipos de información o de análisis para verificar y sustanciar una evaluación, haciendo verificaciones cruzadas de los resultados. La triangulación es útil en el análisis de datos cualitativos. Por ejemplo, considérense los ejemplos siguientes de fuentes mixtas de datos:

- Entrevistas, grupos de discusión y cuestionarios.
- Cuestionarios, datos existentes y paneles de expertos.
- Observaciones, registros de programa y cartografía.
- Entrevistas, diarios y datos existentes.

La combinación de hallazgos de tres fuentes cualesquiera contribuye a corroborar la evidencia de un patrón.

Organizar los datos cualitativos para el análisis

Después de recopilar datos cualitativos, el evaluador tendrá muchas páginas de notas y transcripciones de observaciones, entrevistas y otras fuentes de datos, por eso organizar y sacarle sentido a esta información puede ser problemático.

Se han establecido lineamientos para la correcta organización de los datos, y la documentación de este proceso es importante para demostrar la validez de los hallazgos (CIID, 2008).

Para empezar a organizar la información, los evaluadores:

- Revisan para asegurarse de que los datos estén completos.
- Hacen varias copias de los datos.
- Organizan los datos en diferentes archivos (CIID, 2008).

El evaluador puede organizar los archivos en distintas formas, utilizando copias impresas o archivos electrónicos, algunos de los cuales pueden ser facsímiles obtenidos mediante el escáner. Algunos evaluadores crean cuatro archivos, así: un primer archivo en el que almacenan los datos en orden cronológico; otro en el que mantienen archivos analíticos o notas de bitácora, un tercero en el que mantienen notas pertinentes a la metodología de investigación, y el cuarto, que contiene una copia de todas las notas (CIID, 2008).

001El camino.indd 354 08/11/2010 03:27:55 p.m.

Patton (2002) presenta otras opciones para organizar e informar datos cualitativos, y recuerda a los evaluadores que el análisis de los datos debe partir de las preguntas de evaluación. La elección de la organización de los datos debe representar un esfuerzo por responder las preguntas de evaluación.

- Los enfoques de narrativa presentan los datos cronológicamente (contando la historia de principio a fin), o en retrospectiva (*flashback,* empezando por el final y retrocediendo para describir cómo se llegó al final).
- Los enfoques de casos de estudio presentan información sobre personas o grupos y los incidentes críticos se presentan usualmente en orden de aparición.
- Los marcos analíticos incluyen descripciones de procesos, ilustración de temas clave (a menudo equivalentes a las preguntas primarias de evaluación), organización de preguntas y discusión de conceptos básicos, por ejemplo, liderazgo frente a seguimiento.

Lectura y codificación de datos cualitativos

Los primeros pasos consisten en identificar y utilizar categorías para organizar la información. El paso siguiente es leer los datos completamente. Después de varias lecturas, el evaluador debería empezar viendo temas potenciales. Patton (2002) anota que pensar en los temas más centrales es como construir un índice para un libro o unos rótulos para un sistema de archivos, y sugiere que los evaluadores examinen lo que tienen y le den a cada tema un nombre o un rótulo. Una vez se han organizado en temas los datos, deben codificarse (recuadro 10.1). "Los códigos son dispositivos de rotulación y recuperación eficiente de datos y habilitan y aceleran el análisis" (Miles y Huberman, 1994, p. 65).

La codificación es un proceso iterativo. Aunque es útil crear una lista de códigos antes de iniciar el trabajo de campo, los evaluadores deben recibir, revisar, redefinir, agregar y algunas veces descartar códigos a medida que progresa la evaluación, y las notas de campo sugieren rótulos más guiados de manera empírica.

Hacer el análisis de contenido

El análisis de datos cualitativos se denomina *análisis de contenido,* en el cual se identifica y codifica la presencia de ciertas palabras, frases o conceptos dentro del texto, la conversación y otros medios. Es un enfoque sistemático en el cual se identifican y resumen los mensajes ocultos en los datos.

El análisis de contenido se refiere al análisis de libros, folletos, transcripciones, nuevos informes, otros documentos escritos, conversaciones y medios visuales. Se aplica a narrativas, tales como diarios y bitácoras, e igualmente a las respuestas cualitativas a preguntas abiertas en encuestas, entrevistas y grupos de discusión. El análisis de contenido podría utilizarse, por ejemplo, para examinar los textos de

Análisis de contenido: Tipo de análisis sistemático de datos cualitativos en el que identifica y anota, por medio de códigos, la presencia de ciertas palabras, frases o conceptos dentro del texto, la conversación y otros medios.

Recuadro 10.1 *Ejemplo de codificación*

Los siguientes son ejemplos de códigos establecidos para una evaluación de un programa educativo. La P representa participantes y la S personal.

Código: Ps Re Prog (reacciones de los participantes al programa).
Código: Ps Re Ps (reacciones de los participantes a otros participantes).
Código: Ob PP (observaciones de interacciones de los participantes).
Código: Ob SS (observaciones de interacciones del personal).
Código: Ob SP (observaciones de interacciones entre el personal y los participantes).
Código: Phil (enunciados sobre la filosofía del programa).
Código: Prc (ejemplos de procesos del programa).
Código: P/outs (efectos del programa en los participantes, efectos directos).
Código: S-G (formación de subgrupos).
Código: C! (conflictos).
Código: C-PP (conflictos entre los participantes).
Código: C-SP (conflictos entre el personal y los participantes).
Código: C-SS (conflictos entre el personal).

Si el código se hace manualmente, se escriben las abreviaturas en los márgenes directamente después de los pasajes de datos o de citas pertinentes. Los rótulos completos entre paréntesis son las designaciones para archivos separados que contienen los pasajes con códigos similares.

Fuente: Patton, 2002.

los niños a fin de establecer si cubren el material para el aprendizaje de una materia particular, imparten el material en el nivel de lectura apropiado y lo presentan de manera consistente con el contexto en el cual viven y estudian los niños. Con un análisis más profundo podría examinarse si los textos transmiten una agenda política específica o una interpretación sesgada de la historia.

Análisis de contenido conceptual: Análisis en el que se tiene en cuenta la frecuencia de la ocurrencia de términos seleccionados dentro de un texto.

Análisis del concepto de relación: Análisis que va más allá de determinar la frecuencia para explorar relaciones entre los conceptos identificados.

El análisis de contenido empieza generalmente con la codificación de los datos. En el proceso se asume que las palabras y frases más mencionadas son las que reflejan preocupaciones importantes. Por consiguiente, el análisis de contenido se inicia con frecuencias de palabras, mediciones de espacio (centímetros de columna en el caso de los periódicos), recuentos de tiempo (para tiempo de radio y televisión) y frecuencias de palabras clave.

El análisis de contenido se extiende mucho más allá de la sola contabilización de las palabras. Para analizar el contenido, se codifican y organizan las palabras en categorías manejables; se examinan la frecuencia de respuesta y las relaciones de estas nuevas categorías codificadas (recuadro 10.2).

Recuadro 10.2 *Uso del análisis de contenido para identificar los beneficios percibidos de las bombas manuales*

Para conocer cómo perciben los aldeanos el uso de las bombas manuales para bombear agua, los evaluadores entrevistan a 30 de ellos y analizan los resultados, utilizando el proceso siguiente:

1. Leen una respuesta sí y otra no, y escriben cada respuesta.
2. Escogen las respuestas más frecuentes y las enuncian brevemente. Agrupan cada respuesta principal en una categoría. Si, por ejemplo, las respuestas incluyen "las bombas manuales son mucho más fáciles de usar" y "el agua siempre está limpia y la bomba nunca se ha roto", el evaluador podría identificar las siguientes categorías: facilidad de uso, suministro de agua limpia, confiabilidad.
3. Asegurar que todas las categorías sean mutuamente excluyentes, y que los codificadores puedan identificar las respuestas correspondientes a cada categoría.
4. Completar el procedimiento de codificación.
5. Tabular las frecuencias para cada respuesta.

Fuente: Narayan, 1996.

El análisis de contenido puede dividirse en dos clases: análisis conceptual y análisis relacional. En un *análisis de contenido conceptual* se tiene en cuenta la frecuencia de la ocurrencia de términos seleccionados dentro de un texto. Un *análisis del concepto de relación* va más allá de determinar la frecuencia para explorar relaciones entre los conceptos identificados (Busch y otros, 2005).

Por ejemplo, en un análisis de concepto se podrían identificar categorías, utilizando la lista siguiente:

- Lenguaje compartido sobre el tema (lo que se dio por acordado y lo que requería clarificación de otros participantes).
- Creencias y mitos sobre el tema, compartidos, dados por sentado y cuestionados.
- Argumentos a los que apelan los participantes cuando se cuestionan sus puntos de vista.
- Fuentes de información a las que apelan los participantes para justificar sus puntos de vista y experiencias, y reacciones a ellos de los demás.
- Argumentos, fuentes y tipos de información que estimulan cambios de opinión o reinterpretación de la experiencia.
- Tono de voz, lenguaje corporal y grado de compromiso emocional cuando los participantes conversan entre sí sobre el tema (Catterall y Maclaran, 1997).

Según Kripendorf (2004), el análisis de contenido debe responder seis preguntas:

- ¿Qué datos se analizan?
- ¿Cómo se definen los datos?
- ¿De qué población se tomaron los datos?

- ¿En relación con qué contexto se analizan los datos?
- ¿Cuáles son los límites del análisis?
- ¿Cuál es el objetivo de las inferencias?

Una vez tratadas estas preguntas, pueden hacerse elecciones sobre datos aplicables y no aplicables.

Neuendorf (2006) presenta un diagrama de flujo en el que se muestra el proceso del análisis de contenido (gráfico 10.1). Empieza considerando la teoría y el fundamento para el análisis.

Un buen análisis de contenido es difícil de realizar, y los evaluadores deberían ser conscientes de los problemas potenciales en su aplicación (cuadro 10.2).

Análisis de contenido asistido por computador

Para que puedan utilizarse los computadores en el análisis de contenido, toda la información debe estar en archivos legibles por los computadores. Los evaluadores deben ingresar, pasar por el *escáner* o reproducir los archivos de datos.

Muchos paquetes de *software* ayudan a organizar los datos derivados de observaciones, entrevistas o grupos de discusión; entre ellos están los administradores de bases de datos orientados a texto, procesadores de palabras y *software* de indexación automática. Estos programas se desarrollaron específicamente para el trabajo con aplicaciones de texto. Algunos de los programas de *software* de análisis de contenido pueden utilizarse con otros tipos de medios.

El *análisis cualitativo asistido por computador* (ACAC) –conocido también como *software* de análisis de datos cualitativos (*software* QDAS o QDA)– busca, organiza, clasifica y anota datos textuales y visuales (recuadro 10.3). Los programas de esta clase permiten a los evaluadores visualizar relaciones y elaboraciones teóricas, y les ayudan a formular teorías. Los paquetes incluyen Ethnograph, Qualpro, Hyperqual, Atlas-ti, QSR's N6 (antes NUD*IST) y NVivo 8, AnSWR, HyperRESEARCH, Qualrus y otros.

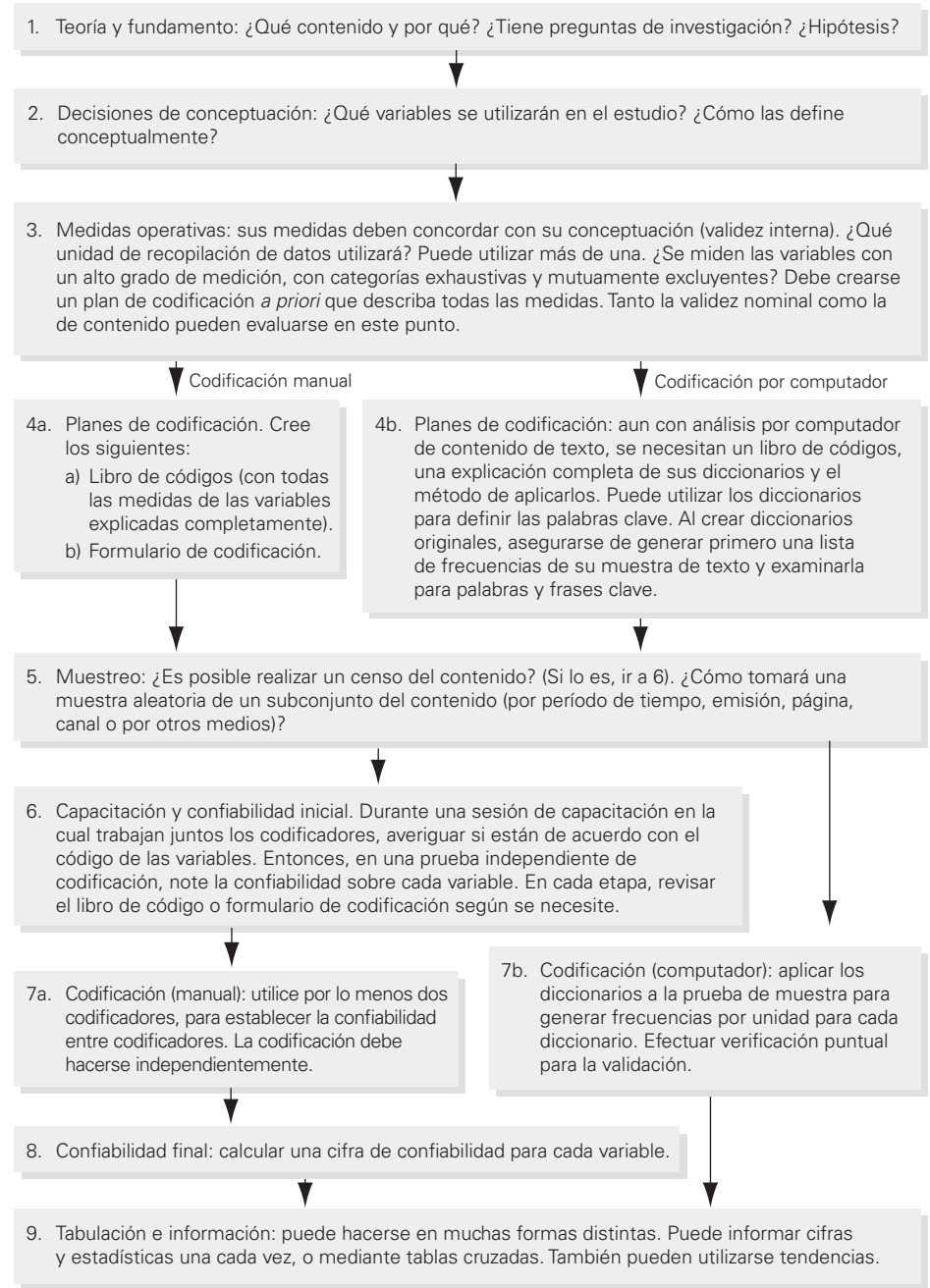
ACAC: Uso del computador para buscar, organizar, categorizar y anotar los datos.

El sitio web de la American Evaluation Association tiene una sección pública sobre recursos. Una de las categorías bajo recursos es la de "qualitative data analysis software" (*software* de análisis de datos cualitativos), en la cual se encuentran descripciones breves de *software* disponible que a menudo incluyen los costos y el alcance o el tipo de medio al que se dirige, como también vínculos a sitios de proveedores de *software*. Las descripciones mencionan también la disponibilidad de descargas o de pruebas gratis.

Análisis manual de datos cualitativos

Porteous, Sheldrick y Stewart (1997) proporcionan valiosas sugerencias para el análisis manual de los datos cualitativos. Por supuesto, si se han recopilado grandes cantidades

Gráfico 10.1 *Diagrama de flujo para investigación de análisis de contenido*



Fuente: Neuendorf, 2006.

Cuadro 10.2 *Fortalezas y problemas potenciales del análisis de contenido*

Fortalezas potenciales	*Problemas potenciales*
Examina directamente la comunicación utilizando textos o transcripciones, y de ahí llega al aspecto central de la interacción social.	Puede consumir demasiado tiempo.
Permite operaciones cuantitativas y cualitativas.	Está sujeto a error, particularmente cuando se utiliza el análisis relacional para obtener un mayor nivel de interpretación.
Proporciona valiosas percepciones históricas y/o culturales sobre el tiempo mediante el análisis de textos.	Está desprovisto a menudo de una base teórica o intentos de obtener inferencias significativas sobre las relaciones y los impactos implicados en un estudio con demasiada liberalidad.
Puede utilizarse para interpretar textos con fines como el desarrollo de sistemas expertos (porque el conocimiento y las reglas pueden codificarse en términos de enunciados explícitos sobre las relaciones entre los conceptos).	Es inherentemente reductor, particularmente cuando trata textos complejos.
Es un medio discreto de analizar interacciones.	Tiende demasiado a menudo a consistir solamente en un conteo de palabras.
Proporciona percepción en modelos complejos de pensamiento humano y uso del lenguaje.	A menudo hace caso omiso del contexto que produjo el texto, y también el estado de cosas después de haberse producido el texto.
	Puede ser difícil de computarizar.

Fuente: Busch y otros, 2005.

Recuadro 10.3 *Estrategias para la codificación electrónica de datos*

Al utilizar ACAC son útiles los consejos siguientes:

- Los memorandos constituyen la forma básica de anotar datos. Pequeñas notas electrónicas permiten a los usuarios adjuntar memorandos a los datos.
- La codificación libre permite a los evaluadores marcar y adjuntar un código a secciones de datos.
- Los procedimientos de codificación automática funcionan de varias maneras, de las cuales la más común es hacer que el programa del computador codifique automáticamente los resultados de la búsqueda. También pueden recodificarse de manera automática los datos con base en búsquedas especificadas anteriormente.
- Las sugerencias de codificación generadas por *software* son una característica novedosa de Qualrus, en el cual un algoritmo sugiere los códigos basado en códigos de aparición anterior.
- N6, HyperRESEARCH y Qualrus ofrecen codificación multimedia. Estos programas permiten codificar secuencias de archivos de audio o video y partes de imágenes. Algunos otros programas ACAC permiten enlaces a archivos multimedia externos.

Fuente: Loughborough University Department of Social Sciences, 2007.

001El camino.indd 360 08/11/2010 03:27:55 p.m.

de datos, se prefiere el *software* asistido por computador. Se empieza teniendo disponibles los materiales siguientes:

- Varios resaltadores (un color diferente para cada pregunta de evaluación).
- Una hoja de trabajo para cada pregunta de evaluación (gráfico 10.2).
- Datos, incluyendo notas, transcripciones y cintas de las entrevistas o grupos de discusión.
- Herramientas de recopilación para cuestionarios, formularios de registro, observaciones o revisiones de gráficos.

Utilizar por lo menos una hoja de trabajo para cada pregunta de evaluación. Escribir cada pregunta en el espacio suministrado en la parte superior de cada hoja de trabajo. Para cada pregunta, escoger un código para identificar los datos. Podría ser el color de una pluma, un lápiz o un resaltador, o un símbolo. Registrar el color o el símbolo en el segundo espacio de la parte superior de cada hoja de trabajo.

Para llenar las hojas, examinar las notas y materiales recopilados hasta el momento y codificar la información, utilizar el procedimiento siguiente:

- Leer las herramientas o notas completas y transcripciones una vez.
- Utilizar resaltadores de diferentes colores para marcar las partes referentes a cada pregunta de evaluación.

Gráfico 10.2 *Hoja de trabajo en blanco para análisis de datos cualitativos*

Tema / Número de referencias	*Citas*	*Hallazgos*

Fuente: Porteous, Sheldrick y Stewart, 1997.

- Regresar y leer cuidadosamente todos los datos pertinentes a la primera pregunta de evaluación.
- En la columna "Tema" de la hoja de trabajo escribir cada opinión, idea o sentimiento pertinente a las expectativas para esa pregunta de evaluación.
- Dejar un espacio entre cada tema, separando espacio para anotar con qué frecuencia aparece cada punto.
- Llevar la cuenta del número de veces que se menciona una opinión, idea o sentimiento.

Tratar el resto de la hoja de trabajo en la forma siguiente (gráfico 10.3):

- De las notas, extraer e insertar las citas que mejor representen cada tema.
- Sacar una conclusión preliminar sobre puntos específicos y escribirlas en la columna "Hallazgos".
- Organizar los hallazgos según tipos o categorías.
- Utilizar números de respuestas (N = x) para dar precisión y sentido de magnitud.

Cuando se utilizan tarjetas de anotaciones al analizar datos cualitativos, la meta es resumir lo que se ha oído o visto en palabras, frases, temas o patrones comunes. Los temas nuevos pueden aparecer después, y habrá que volver a leer el material anterior para verificar si el tema estaba realmente allí desde el principio pero fue omitido, porque en aquel momento no era claro su significado.

Al identificar palabras, temas o patrones, marcar su localización a fin de poder encontrarlos nuevamente si se necesita para verificar citas o contexto. Esto puede ser tedioso la primera vez que se hace, pero con la experiencia es más fácil localizar la información potencialmente importante con mayor rapidez.

En ocasiones, la opinión minoritaria es importante y debe informarse. Utilizar el propio juicio, pero siempre aclarar que solamente uno de unos pocos encuestados expresó esa opinión. Como se anota en el sitio web de la University of the West of England sobre el análisis de datos:

> La vida rara vez viene empacada limpiamente en manojos ordenados. Siempre hay callejones sin salida, temas que se agotan o son inconsistentes entre sí. La tentación en la investigación cualitativa es ignorar las categorías extrañas que no se acomodan limpiamente en la teoría emergente. Estas rarezas son como los calcetines solos que se encuentran en los cajones y de ahí el fenómeno del talego de calcetines. Todos los proyectos de investigación cualitativa tendrán rarezas que desafían la clasificación; en lugar de borrarlos del cuadro deben reconocerse como parte del todo.

Interpretación de los datos cualitativos

Los evaluadores describen datos y los interpretan. Antes de analizar los datos cualitativos, deben presentarse claramente, en forma descriptiva. Interpretar datos significa encontrar posibles vínculos causales, hacer inferencias, adjuntar significados y tratar con casos que contradicen el análisis.

Gráfico 10.3 *Hoja de trabajo llena para análisis de datos cualitativos*

Tema / Número de referencias	*Citas*	*Hallazgos*
Padres 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ////	Narrativa: pienso que el proceso de decidir sería valioso.	Hubo una fuerte sensación de que los padres deberían participar más en la elección de los temas.
Cubrir varios temas por sesión 𝍸 𝍸 𝍸 ///	Algunas veces, apenas llegamos a un tema y ya era tiempo de salir o de pasar a otro asunto.	Muchos participantes (38 de 52 entrevistados) opinaron que debería haber más tiempo para las discusiones.
Tiempo insuficiente empleado en cada tema 𝍸 𝍸 𝍸 ///	Necesitamos más tiempo para discutir.	

Fuente: Porteous, Sheldrick y Stewart, 1997.

Muchas personas temen usar las estadísticas y, en consecuencia, hay una tendencia a pensar que utilizar métodos cualitativos es en alguna forma la opción más fácil. En realidad, un buen análisis cualitativo requiere más de lo que el observador casual ve a primera vista. Aunque el análisis de datos cualitativos es intensivo en mano de obra y consume mucho tiempo, puede revelar percepciones sobre comportamientos o procesos que no pueden obtenerse de datos cuantitativos. Los evaluadores deben planear el tiempo suficiente para realizarlo bien.

Los métodos cualitativos pueden constituir herramientas poderosas para examinar la causalidad. (Un recurso excelente que incluye una guía paso a paso para efectuar análisis de datos cualitativos es Miles y Huberman, 1994). Patton (2002) describe dos clases de análisis cualitativo: inductivo y deductivo. El *análisis inductivo* incluye el descubrimiento de patrones, temas y categorías en los datos. El *análisis deductivo* implica analizar los datos, utilizando un marco existente. Usualmente los datos cualitativos se analizan inductivamente en las primeras etapas (descubriendo categorías, patrones y temas). Una vez establecidas estas categorías, patrones y temas, puede realizarse el análisis deductivo. La fase deductiva incluye poner a prueba y afirmar la autenticidad y lo apropiado del análisis inductivo.

Al trabajar con datos cualitativos existe algún riesgo de sesgo (si no se utiliza el *software* de análisis de contenido); con mucha frecuencia las personas ven lo que desean ver y omiten las cosas que no se conforman a sus expectativas; por eso es una buena ayuda (pero no siempre remedia completamente la situación) contar con otra persona que analice los datos. Al comparar los dos análisis, los evaluadores pueden identificar nuevos temas o hallar distintas formas de entender los datos.

Análisis inductivo: Análisis de datos que incluye el descubrimiento de patrones, temas y categorías.

Análisis deductivo: Análisis de datos, utilizando un marco existente.

Cuando se informan datos cualitativos, no siempre es posible o significativo presentar una cuenta de cuántos o qué porcentaje dijo o hizo algo. Dado que no se hizo la misma pregunta a todos los participantes, es difícil saber cómo se sintió cada uno sobre la pregunta. Otra forma de controlar el sesgo es hacer que dos codificadores revisen los mismos documentos y los codifiquen en cuanto a temas. Si los evaluadores están bien capacitados y las definiciones operativas y los sistemas de clasificación son claros y se han acordado, la tasa de acuerdo sería alta entre ambos evaluadores en su clasificación del material. Un alto nivel de confiabilidad entre quienes clasifican sería un indicador de credibilidad y una baja tasa de confiabilidad indicaría la necesidad de revisar las definiciones operativas, los sistemas de clasificación, o ambos.

En el cuadro 10.3 se resumen las sugerencias para organizar e interpretar datos cualitativos.

Informe de datos cualitativos

En muchas evaluaciones se utilizan datos cualitativos y cuantitativos. Si se utilizan enfoques mixtos de recopilación de datos (cualitativos y cuantitativos), los evaluadores querrán recibir comentarios que clarifiquen e ilustren algunos de los datos cuantitativos. Si, por ejemplo, el 55% de los encuestados estuvo insatisfecho con la accesibilidad de una intervención, es útil incluir una mezcla representativa de comentarios que ilustre las fuentes de la insatisfacción.

El evaluador deseará capturar "citas citables". Deben seleccionarse las declaraciones de los participantes porque presentan claramente un tema o un punto importante que vale la pena enfatizar. En estas palabras existe poder: muchos lectores de los informes recordarán una cita pero no una página de descripción. Para evitar sesgos, los evaluadores deben incluir citas que muestren varios problemas y perspectivas sobre el mismo tema.

Análisis de datos cuantitativos

En el análisis de datos cuantitativos se resume la información numérica recopilada como parte de una evaluación. Los evaluadores ingresan los datos a un archivo de datos en el computador para organizarlos, o utilizan paquetes de *software* con los que se analicen los datos a medida que se ingresan.

Codificar los datos

Los datos cuantitativos deben codificarse para su análisis, pero esto es más fácil que hacer el análisis de contenido de datos cualitativos. La codificación de datos cuantitativos se utiliza cuando se requiere transformar éstos en respuestas numéricas, y permite que los datos se procesen en una forma significativa. Los datos sobre características, como altura, peso, edad y número de días de ausencia al colegio, no requieren codificación, pues ya son numéricos. Otros tipos de datos –por ejemplo, si un encuestado posee cuenta

Cuadro 10.3 *Organizar e interpretar datos cualitativos*

Tarea	*Sugerencia*
Desarrollar las categorías	• Utilizar temas, ideas, palabras y frases recurrentes. • Utilizar categorías suficientemente grandes para capturar una variedad de opiniones, pero no tan grande que no tenga significado. • Crear categorías diferentes entre sí.
Codificar las categorías	• Desarrollar un esquema de codificación. • Desarrollar reglas de decisión exhaustivas y sin ambigüedad para codificar. • Capacitar codificadores para utilizar el esquema de codificación.
Verificar la confiabilidad al utilizar más de un observador	• Efectuar una prueba preliminar con una pequeña muestra de datos cualitativos. • Verificar la confiabilidad entre quienes clasifican. ¿Obtienen los mismos resultados las personas que miden la misma cosa en la misma forma? • Si existen problemas, resolverlos y luego efectuar nuevamente la prueba preliminar.
Analizar los datos	• Poner los datos en orden. • Considerar escribir los datos en tarjetas. • Considerar escribir los datos en una hoja de trabajo. • Considerar utilizar un computador como ayuda para el análisis de datos. • Ordenar los datos para revelar patrones y temas.
Interpretar los datos	• En lo posible, contar con equipos de por lo menos dos personas para revisar y clasificar los datos, con el fin de comparar sus hallazgos y revisarlos si difieren. • Buscar significados en los datos. • Relacionar temas y categorías en los procesos del programa, en sus resultados o en ambos. ¿Predominan más algunos temas cuando los encuestados discuten asuntos de procesos? ¿Son algunos temas más pertinentes cuando los encuestados discuten asuntos de resultados? Buscar explicaciones alternativas y otras formas de comprender los datos.
Compartir y revisar la información	• Compartir información temprano y a menudo con informantes clave. • Hacer que otras personas revisen los primeros borradores con la intención de obtener información, preguntas, otras formas de interpretar los datos y otras posibles fuentes de datos.
Escribir el informe	• Describir los temas principales (enfoque temático), o presentar material cuando refleja lo sucedido a través del tiempo (enfoque de historia natural). • Destacar perspectivas interesantes aun si sólo las han anotado una o dos personas. • Mantenerse enfocado, pues con grandes cantidades de datos es fácil perderse. • Incluir solamente la información importante. Preguntarse si la información responde las preguntas de la evaluación y si será útil para las partes interesadas.

Fuente: adaptado de Porteous, 2005.

bancaria o no– requieren códigos numéricos para su análisis, para lo cual un evaluador puede codificar las respuestas, utilizando un 1 para *sí* y un 2 para *no.*

Otros datos pueden recopilarse como rangos de opinión. Por ejemplo, supóngase que la pregunta es "¿A qué grupo de edad pertenece usted?" Aquí puede asignarse un

código a cada uno de los grupos de edad (menor de 18 = 1, 18-25 = 2, 26-35 = 3, etc.). Algunos consejos son útiles para mantener datos cuantitativos (recuadro 10.4).

Cada persona o registro se denomina un "caso". Los archivos de datos se componen de variables y valores para cada variable. Las variables son cada uno de los ítems que se analizan.

Para los evaluadores, es muy importante indicar cómo se codificó cada variable y a qué categoría se asignó. Los códigos deben ser registrados en un libro de códigos (llamado algunas veces *diccionario de datos*) y los codificadores deben tener acceso al libro de códigos y estar capacitados para usarlo.

Limpieza de los datos

La *limpieza de datos* (llamada también *purificación de datos* o *fregado de datos*) es el proceso de eliminar errores e inconsistencias de los datos con el fin de mejorar su calidad (Rahm y Do, 2000). Los datos con errores o inconsistencias se denominan a menudo "datos sucios". Los analistas de datos estiman que hasta la mitad del tiempo requerido para el análisis se emplea en la limpieza de los datos, y este tiempo usualmente se subestima. Una vez se ha limpiado el conjunto de datos, el análisis es expedito (PANDA, 2000).

Las fuentes comunes de errores en una base de datos incluyen las siguientes:

- Datos faltantes.
- Respuestas como "no aplicable" o en blanco.
- Errores tipográficos o de ingreso de datos.
- Formateo incorrecto de datos.
- Desplazamiento de columnas (datos de la columna de una variable se ingresan en la columna adyacente).
- Fabricación de datos.
- Errores de codificación.
- Errores de medición y de entrevista.
- Datos desactualizados (PANDA, 2000).

Recuadro 10.4 *Consejos para mantener datos cuantitativos*

- Asegurarse siempre de rotular las variables de datos.
- Asegurarse de que el diccionario de datos esté actualizado con respecto a cambios en los rótulos de los datos o códigos de respuestas. Una buena documentación es esencial.
- Generar una copia de respaldo del conjunto de datos para casos de emergencia. Crear conjuntos de datos temporales y permanentes en forma sensata. Pensar en lo que debe hacerse de perderse los datos.
- Mantener siempre una copia del conjunto de datos originales.

Fuente: Child and Adolescent Health Measurement Initiative, 2009.

En algunos casos, los encuestados no saben la respuesta a una pregunta, o se rehúsan a contestarla. En otros casos, un encuestado puede saltarse una pregunta inadvertidamente. Una convención que utilizan muchos evaluadores para estas respuestas es esta:

- No sabe = 8.
- Rehusó contestar o dato faltante = 9.

Por ejemplo, considérese que varias personas han ingresado los datos a una base de datos. Un codificador ingresó todos los datos utilizando mayúsculas, otro los ingresó utilizando mayúsculas solamente para la primera letra del apellido. Un codificador ingresó cada dirección como una entrada, otro ingresó la calle como una entrada, la ciudad como otra y el país como otra. Además, los datos fueron mezclados con otros de una evaluación efectuada 10 años antes, y el nuevo instrumento de la encuesta incluyó más preguntas que las anteriores, de modo que los datos anteriores no incluyen respuesta para muchas preguntas.

Otro problema con los datos es que muchos encuestados se confundieron con algunas preguntas. Las personas que grabaron las respuestas hicieron su mejor esfuerzo sobre cómo tratar estas respuestas, pero cada una utilizó reglas distintas para codificar. En varios casos, los encuestados, a quienes se solicitó seleccionar de una escala de 1 a 5, utilizaron cifras decimales (por ejemplo 2,5). Algunos codificadores redondearon dichas cifras al entero más próximo, otros las redondearon a 2, y otros aún las trataron como datos faltantes.

Todos estos problemas son causados por error humano al responder o al ingresar datos, y muchos de ellos pueden minimizarse si se fijan reglas de codificación de datos para su ingreso y se cumplen estrictamente. Sin embargo, siempre habrá errores en la entrada de datos, y por esta razón debe validarse. Los evaluadores deben fijar reglas para codificar las respuestas y mantener los cuestionarios originales (para referirse a ellos cuando se identifiquen posibles errores), y deben inspeccionar los datos e intentar identificar datos sucios.

Después de ingresados los datos, deben filtrarse y limpiarse antes de analizarlos. Por ejemplo, considérense los datos basados en registros escolares. Para preguntas sobre género, sólo hay dos respuestas posibles: 1 para masculino y 2 para femenino. Si se encuentran otras respuestas, deben considerarse como errores. Para una pregunta sobre la de tener un examen físico, los únicos valores posibles serían 1 para sí, 2 para no, 8 para no sabe y 9 para respuesta faltante o rehusada. Cualquier otra respuesta debe considerarse error de codificación, y para detectar posibles errores en datos sobre altura, el evaluador puede ver las entradas que están muy por encima o por debajo de lo esperado para la edad del estudiante (O'Rourke, 2000b). Si existe preocupación sobre la exactitud, el evaluador puede regresar a las hojas originales para verificar la exactitud (o inexactitud) de los datos codificados. (Chapman [2005] ofrece una introducción útil sobre principios y métodos de limpieza de datos,

Limpieza de los datos: Es el proceso de eliminar errores e inconsistencias de los datos con el fin de mejorar su calidad.

y ofrece como referencia varios recursos en línea y *software* disponible. Aunque trata específicamente temas de biodiversidad, gran parte de la información suministrada es de uso general.) Para demostrar técnicas de limpieza de datos, Cody [s.f.] ofrece un ejercicio, cuyo vínculo se encuentra al final de este capítulo.

Hay programas de *software* de computador disponibles para la limpieza de datos. Los programas revisan datos que se encuentran fuera de los rangos. Un ejemplo de éstos es WinPure. Es crucial mantener un registro de las decisiones tomadas sobre la limpieza de datos.

Uso de la estadística

Los datos cuantitativos se analizan utilizando la estadística. Esta sección introduce algunos de los conceptos estadísticos más importantes que deben conocer las personas que realicen o lean evaluaciones para el desarrollo.

Los estadísticos dividen la estadística en dos grandes categorías:

- *Estadística descriptiva,* que (en su definición más estrecha) describe y resume datos cuantitativos.
- *Estadística inferencial,* que se utiliza usualmente para analizar datos de muestras aleatorias, pronosticando un rango de valores de población para una variable cuantitativa o cualitativa, con base en información de la muestra aleatoria para esa variable. Parte de la predicción incluye una declaración de confiabilidad, que expresa la probabilidad de que el verdadero valor de la población se encuentre dentro de un rango de valores especificado.

Estadística descriptiva

Usualmente, los datos se resumen utilizando dos clases de estadística descriptiva:

- *Medidas de tendencia central:* forma de describir un grupo de datos para indicar el punto central.
- *Medidas de dispersión:* forma de describir un grupo de datos para indicar cuán dispersos se encuentran los datos.

Estadística descriptiva: Estadística utilizada para describir y resumir datos cuantitativos.

Estadística inferencial: Estadística utilizada para hacer inferencias sobre una población a partir de una muestra.

Medidas de tendencia central: Medidas estadísticas que destacan la similitud de las características.

Medidas de dispersión: Medidas estadísticas que destacan las diferencias de las características.

Medidas de tendencia central. Las tres medidas de tendencia central se denominan algunas veces las tres M.

Un grupo de datos se arregla a menudo de un modo gráfico llamado distribución de frecuencias, para resumirlo de tal forma que se indica la frecuencia de cada variable del grupo. El valor central se determina entonces utilizando una o más de las tres medidas de tendencia central: la media, la mediana y la moda.

Moda: es la respuesta más frecuente.
Mediana: es el punto o el valor que se encuentra en la mitad de una distribución; la mitad de todos los valores de la distribución son mayores que él y la mitad son menores. En conjuntos de datos de número par, se define la mediana como el promedio de los dos casos de la mitad.
Media: es la suma de todos los valores recopilados dividida entre el número de valores (tamaño de la muestra).

Las dos estadísticas más utilizadas son la media y la mediana.

En el cuadro 10.4 se presentan datos sobre la proporción de población urbana en 16 países. Supóngase que se desea resumir esta información e informar el urbanismo medio (promedio) para este grupo de países. La media sería la suma de las proporciones urbanas dividida entre el número de países: (90 + 64 + 84 + … + 93)/16 = 1.141/16 = 71,3. Los dos casos de la mitad por encima y por debajo de los cuales está el 50% de todos los casos son 71 y 73, por lo que la mediana es 72 (73 + 71 = 144/2 = 72). La moda es 47,6. Nótese que en este caso la media y la mediana son aproximadamente iguales, pero la moda es muy diferente.

La medida de tendencia central que se ha de utilizar depende del tipo de los datos: nominal, ordinal o intervalo/razón (cuadro 10.5).

- *Datos nominales.* Algunas veces llamados *datos categóricos*, son datos que se ajustan a una de varias categorías que no se traslapan, como género, religión o país de origen.
- *Datos ordinales.* Son datos que pueden colocarse en una escala ordenada, pero la "distancia" entre respuestas consecutivas no es necesariamente la misma. Las

Cuadro 10.4 *Porcentaje de población urbana por países de América Latina y América Central*

País	*Porcentaje de la población que vivía en zonas urbanas en 2007*	*País*	*Porcentaje de la población que vivía en zonas urbanas en 2007*
Argentina	90	Honduras	47
Bolivia	64	México	76
Brasil	84	Nicaragua	59
Chile	88	Panamá	71
Colombia	73	Paraguay	59
Costa Rica	62	Perú	73
Ecuador	63	Uruguay	92
Guatemala	47	Venezuela	93

Fuente: Banco Mundial. Indicadores de Desarrollo Mundial, 2008, 162-64.

Cuadro 10.5 Medidas de tendencia central preferidas según el tipo de datos

Tipo de datos	*Mejor medida de tendencia central*
Nominal	Moda
Ordinal	Moda o mediana
Intervalo/razón	Moda, mediana o media

Fuente: los autores.

escalas que van de "más importante" a "menos importante" o "fuertemente de acuerdo" a "fuertemente en desacuerdo" ilustran datos ordinales. Los datos ordinales carecen de un punto cero.

- *Datos de intervalo/razón.* Son números reales. Estos datos tienen un punto cero e intervalos fijos, como una regla, que pueden dividirse y compararse con otros números de la razón.

Para los datos de intervalo/razón, la elección dependerá también de la distribución. Si ésta tiene forma de campana, la media, la mediana y la moda deben estar muy cerca entre sí. En este caso, la media sería la mejor medida de tendencia central. En contraste, si los puntajes incluyen unos pocos muy altos o unos pocos muy bajos, la media ya no estará cerca del centro. En esta situación, la mediana sería una mejor medida de la tendencia central.

Medidas de dispersión. Dos medidas se utilizan usualmente para medir la dispersión de las variables cuantitativas: el rango y la desviación estándar. El *rango* es la diferencia entre el valor más alto y el más bajo de una variable. Utilizando los datos del cuadro 10.4, el *rango* para los porcentajes de población urbana es 93 – 47 = 46.

Datos nominales: Datos que se ajustan a una de varias categorías que no se traslapan, como género, religión o país de origen.

Datos ordinales: Datos que pueden colocarse en una escala ordenada.

Datos de intervalo/razón: Datos con un punto cero e intervalos fijos, como una regla, que pueden dividirse y compararse con otros números de la razón.

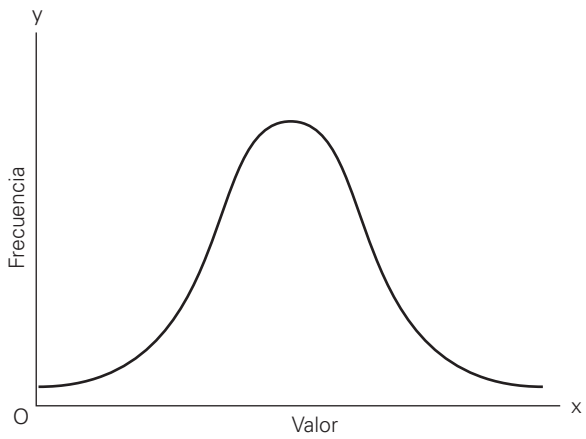
Rango: Diferencia entre el valor más alto y más bajo de una variable.

El rango no es muy revelador porque solamente está determinado por dos observaciones, y se ignora el resto de los casos. Cuando los dos valores finales son extremos, el rango no ofrece orientación sobre dónde caen los demás valores.

La medida de dispersión más utilizada para datos de intervalo/razón es la desviación estándar, la cual es una medida de la dispersión de los valores a cada lado de la media. Cuanto más difieran los valores de la media, mayor será la desviación estándar.

Para comprender mejor la noción de la desviación estándar, es importante comprender la *distribución normal* (gráfico 10.4), llamada algunas veces *curva de campana* debido a su forma. En una distribución normal, la mayor parte de los

Gráfico 10.4 *La distribución normal*



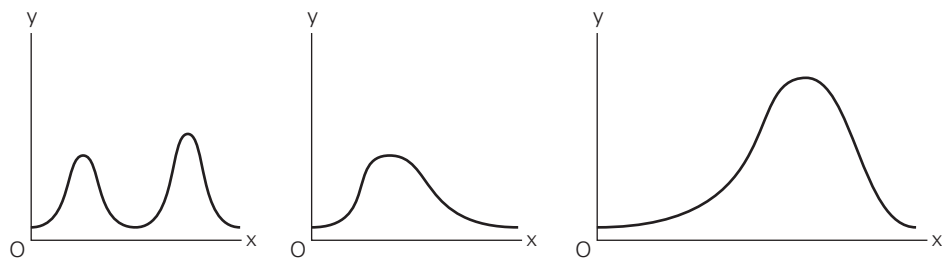
Fuente: los autores.

datos se encuentra hacia la mitad de la distribución, y cada vez hay menos valores en cada extremo de la distribución.

Los datos no siempre presentan una distribución normal y algunas distribuciones tienen curvas más planas, otras más inclinadas, o curvas que suben en un extremo o el otro (gráfico 10.5).

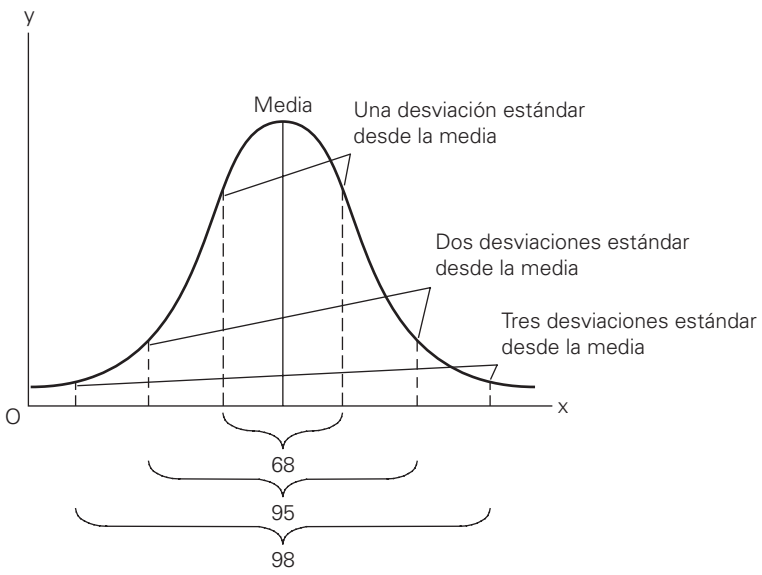
La desviación estándar mide la concentración con que se agrupan los datos alrededor de la media. Mide la distancia desde la media. En una distribución normal, una desviación estándar desde la media en cualquier dirección sobre el eje horizontal representa aproximadamente el 68% de los datos (gráfico 10.6). Dos desviaciones estándar desde la media representan aproximadamente el 95% de los datos y tres desviaciones estándar desde la media representan aproximadamente un 98% de los datos.

Gráfico 10.5 *Distribuciones no normales*



Fuente: los autores.

Gráfico 10.6 *Desviaciones estándar en una distribución normal*



Fuente: los autores.

Si la curva de un conjunto de datos fuese más plana, la desviación estándar sería mayor. El valor de la desviación estándar indica cuán dispersos están los datos con respecto a la media (recuadro 10.5).

Si todos obtienen 75 puntos en una prueba, la media sería 75 y la desviación estándar sería 0. Si todos obtienen un puntaje entre 75 y 80, con un puntaje promedio de 75, la desviación estándar sería más pequeña que si todos obtuviesen puntajes entre 40 y 90 (con una media de 75). Dicho de otra forma:

Desviación estándar pequeña = no mucha dispersión

Desviación estándar grande = mucha dispersión

La desviación estándar es superior al rango porque permite que cada caso tenga impacto en su valor.

Medidas descriptivas utilizadas con frecuencia. Algunas veces se hace una pregunta de cuenta específica ("¿Cuántas cabras posee usted? ¿Con qué frecuencia utiliza el banco?") Las respuestas a tales preguntas se informan en términos absolutos y en términos de porcentajes.

Recuadro 10.5 *Cálculo de la desviación estándar*

La desviación estándar de una distribución se calcula como sigue:

1. Se calcula la media de los datos.
2. Se resta la media de cada punto de datos para encontrar la desviación.
3. Se eleva al cuadrado cada desviación.
4. Se suman todos los cuadrados de las desviaciones.
5. Se resta 1 del número de puntos de datos.
6. Se divide la suma de los cuadrados de las desviaciones entre el resultado del paso 5 (número de casos en la lista menos 1).
7. Se saca la raíz cuadrada al resultado del paso 6.

   La fórmula para calcular la desviación estándar es así:

$$\sigma = \sqrt{\frac{\Sigma\,(x - \bar{x})}{N - 1}}$$

donde σ = desviación estándar, Σ = sumatoria de, y $\bar{x}$ = la media.

Aun con una muestra pequeña toma mucho tiempo calcular la desviación estándar. Afortunadamente, la mayoría de los programas estadísticos, incluyendo Excel y SPSS para Windows, pueden efectuar los cálculos.

Otras veces se solicita a las personas dar opiniones a lo largo de una escala. Por ejemplo, un evaluador podría preguntar si los encuestados podrían aplicar lo que aprendieron, ofreciéndoles una escala de cinco puntos desde "definitivamente no" hasta "mucho". Al analizar esta clase de datos, establecer una regla de decisión: concentrarse en el porcentaje que respondió en los extremos de la escala, en aquellos que respondieron a cada lado de la categoría de la mitad, o en la respuesta promedio. Aunque no hay reglas firmes aquí, algunos lineamientos pueden ser de utilidad (recuadro 10.6).

Considérese una encuesta de clientes de un centro de salud (cuadro 10.6). Una forma de analizar estos datos es informar que la mitad de los encuestados están de acuerdo o muy de acuerdo en que reciben un buen servicio de salud, y un 55% está de acuerdo o muy de acuerdo en que el personal de la clínica está dispuesto a responder preguntas. Sin embargo, un 60% está de acuerdo o muy de acuerdo en que esperan mucho tiempo antes de ser atendidos. En este análisis, el evaluador decidió informar los porcentajes combinados de acuerdo y muy de acuerdo.

Si los datos fuesen diferentes, el evaluador podría utilizar una estrategia diferente. Considérense, por ejemplo, los resultados presentados en el cuadro 10.7.

En este caso en el análisis se podría anotar que el 80% de los encuestados está de acuerdo o muy de acuerdo en que reciben buen servicio de salud, y el 70% está de acuerdo o muy de acuerdo en que esperan mucho tiempo para ser atendidos. La mayor fortaleza parece ser la disposición del personal para responder preguntas, con el 95% de

Recuadro 10.6 *Lineamientos para analizar datos cuantitativos de encuestas*

Los lineamientos siguientes pueden ayudar a los evaluadores a analizar datos cuantitativos de encuestas:

1. Seleccionar una forma estándar para analizar los datos y aplicarla en forma consistente.
2. No combinar la categoría de la mitad con las categorías de cada extremo de la escala.
3. No informar una categoría de "acuerdo" o "desacuerdo" sin informar también las categorías "muy de acuerdo" y "muy en desacuerdo" (si se utilizan).
4. Analizar e informar tanto porcentajes como cifras.
5. Informar el número de encuestados como punto de referencia.
6. Si existe poca diferencia en los resultados, aumentar la marca de referencia: ¿a qué se asemejan los resultados cuando el foco se sitúa en las preguntas con una mayoría de respuestas "muy satisfecho" o "muy en desacuerdo"?
7. Recordar que el análisis de datos es un arte y una destreza; se hace más fácil con capacitación y práctica.

Fuente: los autores.

Cuadro 10.6 *Opiniones de clientes sobre servicios de salud en la clínica local (1) (porcentaje de encuestados)*

1. Considere sus experiencias con la clínica de salud local. ¿Está usted de acuerdo o en desacuerdo con las afirmaciones siguientes?

Afirmación	*Muy en desacuerdo*	*En desacuerdo*	*Ni de acuerdo ni en desacuerdo*	*De acuerdo*	*Muy de acuerdo*
Espero largo tiempo antes de ser atendido.	10	20	10	35	25
El personal está dispuesto a responder mis preguntas.	5	10	30	30	25
Recibo un buen servicio de salud en la clínica.	15	25	10	25	25

Fuente: los autores.
Nota: N = 36.

los encuestados que informan que están de acuerdo o muy de acuerdo con que reciben un buen servicio.

Describir dos variables al mismo tiempo. A veces, los evaluadores desean describir dos variables al mismo tiempo. Por ejemplo, supóngase que desean describir la composición

de clases prácticas y clases teóricas. Para cada clase, desean saber el porcentaje de muchachos y el de muchachas. El análisis de los datos muestra que a las clases prácticas acude el 55% de los muchachos y el 45% de las muchachas, mientras a las teóricas tradicionales acude el 55% de las muchachas y el 45% de los muchachos.

Tabulación cruzada: Distribución conjunta de dos o más variables, usualmente presentada en una tabla.

Una *tabulación cruzada* (o "tab-cruz") presenta la distribución conjunta de dos o más variables, habitualmente en forma de matriz. Mientras una distribución de frecuencias presenta la distribución de una variable, una tabla de contingencia describe la distribución de dos o más variables simultáneamente. Cada celda muestra el porcentaje y el número de encuestados que da una combinación específica de respuestas.

Los datos sobre inscripción de las clases pueden interpretarse así: en esta muestra, los muchachos tienen una probabilidad algo mayor (55%) que las muchachas (45%) de tomar las clases prácticas. Este hallazgo sugiere que existe una relación entre el género y la inscripción a esas clases. Pero, ¿qué tan fuerte es? Al observar las medidas de asociación, es importante comprender los conceptos de variables independientes y variables dependientes.

Las *variables independientes* son las que explican un cambio en otra variable (dependiente). Por ejemplo, en un curso de capacitación las variables independientes podrían incluir la experiencia de los instructores, los antecedentes de los participantes, el plan de estudios, la duración de la capacitación, la modalidad utilizada, etc.

Las *variables dependientes* son aquellas cuya explicación se busca. Para un curso de capacitación, las variables dependientes podrían ser los puntajes en una prueba de conocimiento, las notas en una matriz de diseño para una evaluación, o el mejoramiento de los diseños de evaluación.

Variable independiente: Variable que explica un cambio en otra variable (dependiente).

Variable dependiente: Variable cuya explicación se busca.

Cuadro 10.7 Opiniones de clientes sobre servicios de salud en la clínica local (2) (porcentaje de encuestados)

1. Considere sus experiencias con la clínica de salud local. ¿Está usted de acuerdo o en desacuerdo con las siguientes afirmaciones?

Afirmación	*Muy en desacuerdo*	*En desacuerdo*	*Ni de acuerdo ni en desacuerdo*	*De acuerdo*	*Muy de acuerdo*
Espero largo tiempo antes de ser atendido.	50	20	10	15	5
El personal está dispuesto a responder mis preguntas.	0	5	0	30	65
Recibo un buen servicio de salud en la clínica.	0	20	0	55	25

Fuente: los autores.
Nota: N = 36.

A menudo, los evaluadores están interesados en si existe una diferencia en los valores promedio de una variable cuantitativa para un par de muestras. Por ejemplo, podrían estar interesados en investigar preguntas como las siguientes:

- ¿La producción promedio de las cosechas es mayor después de un proyecto de irrigación que antes del proyecto?
- ¿Las encuestas de pacientes de un hospital viejo y de pacientes de un hospital construido por un proyecto de desarrollo, revelan diferencias en el porcentaje de pacientes satisfechos con el servicio?

Los evaluadores deben determinar si las diferencias aparentes indican diferencias reales en las medias de las poblaciones, o una variación aleatoria en las muestras tomadas para las dos poblaciones. En una prueba estadística, se supone usualmente que no existe diferencia entre las dos medias de las poblaciones (o proporciones). Este tema se trata en la sección siguiente sobre estadística inferencial.

Las *medidas de asociación* (o relación) indican la fuerza de la relación entre las variables. La simple asociación no demuestra la causa. Sólo puede generalmente sugerir la posibilidad de una relación causal si la medida de asociación es fuerte.

Las medidas de asociación se informan usualmente en términos de una escala que va de -1 a +1. Una medida con signo positivo significa que las variables cambian en la misma dirección: ambas suben o bajan. Esto se llama una *relación directa.* Una relación positiva perfecta tendría un valor de +1.

Medidas de asociación: Medias que indican la fuerza de la relación entre variables.

Relación directa: Relación en la que las variables se mueven en la misma dirección (ambas hacia arriba o hacia abajo).

Relación inversa: Relación en la que las variables se mueven en direcciones opuestas.

Un signo negativo indica que las variables tienen una *relación inversa,* lo que significa que se mueven en direcciones opuestas (por ejemplo, al aumentar la edad, la condición de salud disminuye). Una relación negativa perfecta tendría un valor de -1. Cuanto más cercana a cero es la medida, más débil es la relación (una relación que no muestra ninguna asociación tendría un valor de 0); cuanto más cercana a +1 o a -1 la medida, la relación es más fuerte.

Estadística inferencial

La estadística inferencial permite a los evaluadores calcular un estimado sobre una población, basado en una muestra aleatoria seleccionada de esa población. El mayor temor al usar una muestra aleatoria es que los resultados reflejen algún capricho de la muestra y, por consiguiente, no revelen una imagen precisa de la población. Si los evaluadores hubiesen seleccionado una muestra diferente, ¿serían sus resultados razonablemente similares o bastante diferentes?

Las pruebas de significación estadística miden la probabilidad de obtener los mismos resultados si realmente no hubiese diferencia en la población total. Los evaluadores llaman a esto la hipótesis nula, y siempre está basada en una diferencia de cero en la población.

Supóngase que una encuesta basada en una muestra aleatoria de personas en Pakistán muestra que hubo una diferencia de 5.000 rupias en el ingreso anual de hombres y mujeres. La prueba podría expresarse de esta forma: si realmente no hay diferencia en la población, ¿cuál es la probabilidad de encontrar una diferencia de 5.000 rupias en una muestra aleatoria? Si hay una probabilidad del 5% (0,05) o menos, concluimos que los resultados de la muestra constituyen un estimado preciso de la población, que hay una diferencia de unas 5.000 rupias, y que la diferencia es estadísticamente significativa.

Un valor *p* (probabilidad de error) de 5% significa que el evaluador está 95% seguro de que los resultados de la muestra no se deben al azar. Significa que los resultados son estadísticamente significativos al nivel del 0,05.

Todas las pruebas de significación estadística se basan parcialmente en el tamaño de la muestra. Si la muestra es muy grande, es probable que las diferencias pequeñas sean estadísticamente significativas. Los evaluadores deben decidir si las diferencias son importantes, dada la naturaleza de su investigación. Tres pruebas estadísticas utilizadas a menudo son la prueba de chi cuadrado, la prueba *t* y la prueba del análisis de variancia. A continuación se describe brevemente cada una.

Prueba de chi cuadrado. Pese a que no es la medida más fuerte de una asociación, la *prueba de chi cuadrado* es una de las pruebas estadísticas más populares por ser fácil de calcular y de interpretar. El propósito de la prueba de chi cuadrado es determinar si las frecuencias observadas difieren marcadamente de las frecuencias que se esperan por azar. Esta prueba se utiliza para comparar dos valores nominales (por ejemplo, estado marital y afiliación religiosa). Se utiliza también para comparar dos variables ordinales (respuestas en escalas), o una combinación de variables nominales y ordinales.

Prueba de chi cuadrado: Prueba estadística utilizada para determinar si las frecuencias observadas difieren marcadamente de las frecuencias que se esperan por azar.

El estadístico chi cuadrado es la suma de la contribución de cada una de las celdas individuales en una tabla de datos. Cada celda de la tabla contribuye algo al estadístico chi cuadrado general. Si una celda dada difiere marcadamente de la frecuencia esperada, la contribución de esa celda al estadístico chi cuadrado general es grande. Si una celda se encuentra cerca de la frecuencia esperada para esa celda, su contribución al estadístico chi cuadrado es baja. Un estadístico chi cuadrado grande indica que –en algún lugar en la tabla– las frecuencias observadas difieren marcadamente de las frecuencias esperadas. No dice cuál(es) celda(s) está(n) causando que el chi cuadrado sea grande, sólo que está(n) allí. El chi cuadrado mide si dos variables son independientes entre sí con base en los datos observados.

La prueba de chi cuadrado mide la significación de las tabulaciones cruzadas. Los valores de chi cuadrado no deben calcularse para porcentajes. Las tabulaciones cruzadas deben convertirse de vuelta a contabilizaciones absolutas (números) antes de efectuar la prueba. La prueba de chi cuadrado es problemática también cuando cualquier celda tiene una frecuencia de menos de cinco. (Para una discusión profunda de este tema, *ver* Fienberg, 1980).

001El camino.indd 377 08/11/2010 03:27:56 p.m.

Prueba *t*: Prueba estadística utilizada para determinar si un grupo de valores numéricos es estadísticamente más alto o más bajo que otro.

Prueba t. Al observar las diferencias entre los valores para dos grupos, los evaluadores deben juzgar la diferencia entre sus medias relativa a la dispersión o la variabilidad de sus valores. Una *prueba t* hace exactamente eso. Se utiliza para determinar si un grupo de valores numéricos es estadísticamente más alto o más bajo que otro grupo de valores. Este análisis es apropiado cuando se comparan medias de dos grupos. Los evaluadores utilizan los estadísticos *t* para comparar los valores promedio de un grupo afectado por un proyecto con los valores promedio de un grupo de control.

Análisis de la variancia (ANOVA). La prueba *t* es muy engorrosa para utilizarla cuando se van a comparar tres o más grupos. Cuando en una evaluación es necesario comparar las medias de varios grupos diferentes a la vez, es mejor utilizar el análisis de variancia, ANOVA.

ANOVA es una técnica estadística para evaluar diferencias entre conjuntos de datos. Utilizando Excel, se pueden hacer comparaciones simultáneas entre dos o más medias para determinar si el cambio observado se debe a una acción aleatoria, o probablemente a un factor, o a una combinación de factores puestos a prueba. El ANOVA evalúa así cómo las variables independientes nominales influyen en una variable dependiente continua. Asume que las poblaciones para todos los grupos que se comparan tienen desviaciones estándar iguales (supuesto de la homogeneidad de variancia), y que las muestras se seleccionan de la población en forma aleatoria. Es importante verificar que se mantienen estos supuestos antes de utilizar el ANOVA. Las pruebas en un ANOVA se basan en la razón F: la variación debida a un tratamiento experimental o efecto dividida entre la variación debida a error experimental. La hipótesis nula es que esta razón es igual a 1,0 (esto es, el efecto de tratamiento es el mismo que el error experimental). Esta hipótesis se rechaza si la razón F es lo bastante grande para que la posibilidad de que sea igual a 1,0 sea más pequeña que algún nivel asignado de antemano, tal como 0,05 (1 de cada 20).

Análisis de la variancia: Técnica estadística para evaluar diferencias entre conjuntos de datos.

Vínculo de datos cuantitativos y cualitativos

Miles y Huberman (1994) discuten cómo pueden vincularse los datos cualitativos y los cuantitativos. Empiezan su discusión con una cita de Fred Kerlinger, un investigador cuantitativo muy renombrado: "No existen los tales datos cualitativos. Todo es o 1 o 0". Ofrecen entonces una opinión opuesta: todos los datos son básicamente cualitativos.

El debate sobre datos cuantitativos frente a cualitativos se ha realizado con furia durante muchos años en los Estados Unidos, pero no ha constituido en general un problema en las cuestiones sobre el desarrollo. En la evaluación para el desarrollo se utilizan datos cuantitativos y cualitativos para comprender el mundo: "Las cantidades son de cualidades y la medida de una cualidad es sólo la magnitud expresada en su me-

dida" (Miles y Huberman, 1994). Los métodos cualitativos proporcionan más contexto; los enfoques cuantitativos permiten generalización de los hallazgos a otras situaciones.

Miles y Huberman anotan que vincular datos cuantitativos y cualitativos en un diseño de estudio permite a los evaluadores:

- Confirmar o corroborar cada clase de datos por vía de la triangulación.
- Elaborar o desarrollar análisis, permitiendo un detalle más valioso.
- Indicar nuevas líneas de pensamiento por medio de la atención a sorpresas o paradojas, "volteando las ideas" y proporcionando una perspectiva fresca.

Greene, Caracelli y Graham (1997) describen el valor epistemológico y político de mezclar métodos en evaluación:

- *Epistemológico*: puede conocerse mejor algo si se utilizan múltiples formas de conocer para referirse a eso.
- *Político*: todas las formas de conocimiento son parciales; por tanto, deben valorarse y respetarse las múltiples y diversas formas de conocer.

Ellos anotan que "una buena evaluación con métodos mixtos invita activamente a diversas formas de pensamiento y valoración para trabajar en concierto por un mejor entendimiento", y "diferentes clases de métodos están mejor adecuados al aprendizaje sobre diferentes clases de fenómenos".

Hawkins (2005) relaciona los siguientes beneficios de utilizar un enfoque integrado de métodos mixtos en la evaluación:

- Pueden incorporarse revisiones de consistencia con procedimientos de triangulación que permitan calcular dos o más estimados independientes para variables clave.
- Pueden obtenerse diferentes perspectivas. Por ejemplo, mientras los evaluadores consideran que el ingreso o el consumo son los indicadores clave del bienestar familiar, los casos de estudio pueden revelar que las mujeres se preocupan más por la vulnerabilidad, la impotencia o la exposición a la violencia.
- El análisis puede realizarse en diferentes niveles. Los métodos de encuestas pueden proporcionar buenos estimados de bienestar a nivel individual, familiar o comunitario, pero son mucho menos eficaces para analizar procesos sociales (por ejemplo, conflicto social), o realizar análisis institucional (por ejemplo, cuán eficazmente operan los servicios públicos, o cómo la comunidad percibe su operación). Muchos métodos cualitativos están diseñados para analizar asuntos como procesos sociales, comportamiento institucional, estructura social y conflicto.
- Pueden ofrecerse oportunidades para retroalimentación que ayude a los evaluadores a interpretar los hallazgos. Los informes de encuestas incluyen frecuentemente referencias a inconsistencias aparentes en los hallazgos, o diferencias interesantes entre grupos que no pueden explicarse con el análisis de los datos. En la mayoría

de las evaluaciones cuantitativas, una vez finalizada la fase de recopilación de datos, no es posible volver al campo para recopilar datos adicionales.
- Los evaluadores de encuestas frecuentemente se refieren al uso de métodos cualitativos para buscar casos extremos. En muchos casos, el analista tiene que tomar una decisión arbitraria en cuanto a si debería excluirse a un encuestado que informa condiciones significativamente por encima o por debajo de la norma.
- Los beneficios percibidos de los enfoques integrados dependen de los antecedentes del evaluador. Desde la perspectiva del evaluador cuantitativo, un componente cualitativo ayudará a identificar problemas clave que deben atenderse en la evaluación; refinar las preguntas para emparejar las opiniones de los encuestados, y suministrar información sobre el contexto social, económico y político dentro del cual tiene lugar la evaluación. Es posible también regresar al campo para hacer seguimiento a los hallazgos interesantes.
- Los evaluadores cualitativos pueden beneficiarse con el uso de métodos cuantitativos. Pueden utilizarse métodos de muestreo para generalizar los hallazgos a la población en general. La selección de muestras puede coordinarse con encuestas en curso o anteriores, de manera que los hallazgos del trabajo cualitativo puedan compararse con los hallazgos de la encuesta. El análisis estadístico puede utilizarse para controlar características familiares y las condiciones socioeconómicas de diferentes zonas de estudio, eliminando así explicaciones alternativas de los cambios observados.

Hawkins (2005) discute cuándo emplear y cuándo no un enfoque de métodos mixtos. De acuerdo con él, los métodos mixtos deberían emplearse en las circunstancias siguientes:

- Se requiere un entendimiento profundo de la intervención y su contexto.
- El presupuesto es limitado, existen restricciones de tiempo, y la triangulación puede ayudar a validar información recopilada en diferentes fuentes utilizando distintos métodos y muestras pequeñas.

Los métodos mixtos no deberían emplearse en las circunstancias siguientes:

- Las preguntas pueden responderse utilizando un solo método.
- La generalización de los hallazgos es necesaria y los indicadores o las medidas son expeditos.
- Los evaluadores con experiencia en los métodos elegidos no están disponibles para el estudio completo (es mejor utilizar sólo aquellos métodos para los cuales se tiene disponibilidad de pericia).
- El compromiso de los interesados principales con un paradigma particular, excluyendo todos los demás, es tan fuerte que continuarán sin convencerse de un enfoque de métodos mixtos sin importar lo bueno que éste sea.
- El tiempo disponible en la etapa de análisis e interpretación es muy limitado.

001El camino.indd 380 08/11/2010 03:27:56 p.m.

Resumen

El análisis de datos cualitativos se utiliza para datos no numéricos. Los datos cualitativos pueden recopilarse, por ejemplo, utilizando observaciones no estructuradas, entrevistas abiertas, análisis de documentos escritos y transcripciones de grupos de discusión. Las notas tomadas durante la recopilación de datos cualitativos son muy importantes, por lo que deben ser detalladas.

El análisis de contenido es un proceso para analizar datos cualitativos. El análisis de datos cualitativos es intensivo en mano de obra y consume mucho tiempo, pero puede revelar información valiosa.

Después de recopilar datos cualitativos, los evaluadores deben organizarlos. Los datos pueden ordenarse de tal modo que aparezcan patrones y características comunes. Una vez ordenados los datos (manual o electrónicamente) pueden clasificarse para luego interpretarse.

Los datos cuantitativos se analizan utilizando estadísticas descriptiva e inferencial e inferencial. La descriptiva resume los datos y describe el valor central por medio de las tres M: media, mediana y moda. Las medidas comunes de dispersión son el rango y la desviación estándar. Otros estadísticos descriptivos utilizados habitualmente incluyen la distribución de frecuencias, los porcentajes, las tasas, razones y tasas de cambio.

La estadística inferencial permite a los evaluadores hacer estimados sobre una población con base en una muestra aleatoria seleccionada de esa población. Las herramientas comunes de estadística inferencial son la prueba de chi cuadrado, la prueba *t* y el análisis de variancia.

Los evaluadores utilizan usualmente métodos cualitativos y cuantitativos. Utilizar más de un método tiene muchos beneficios en muchos casos. En los casos en que solamente se plantean pocas preguntas fáciles de responder, suele recomendarse un solo enfoque.

001El camino.indd 381 08/11/2010 03:27:56 p.m.

Actividades del capítulo 10

Ejercicio de aplicación 10.1: codificar y analizar datos cualitativos

Recopile varios artículos largos de periódicos sobre un tema de desarrollo. Utilizando esos artículos configure una tabla o una red escrita a mano, rotulando las columnas "artículo" y "extracto". Luego ingrese extractos de los artículos que encuentre importantes. Al identificar temas de los artículos, ingréselos como encabezados de columnas adicionales. Luego marque la celda apropiada en donde un extracto contenga cada tema. Escriba una narración que resuma los hallazgos de los artículos.

Ejercicio de aplicación 10.2: evitar errores comunes al interpretar datos cuantitativos

1. El 80% de los encuestados indica que encontraron útil un programa. ¿Cuál sería la mejor forma de informar los hallazgos: "el programa es útil" o "los participantes encontraron útil el programa"? Se solicitó a los encuestados identificar las barreras y apoyos al programa. ¿Cuál es el problema con informar los resultados en términos de pros y contras?
2. En una encuesta se solicitó a estudiantes calificar varios componentes de un curso. La mayoría de los estudiantes calificó cada uno de los componentes en forma positiva. ¿Cuál es el problema con informar que la mayoría (el 70%) de los estudiantes consideró que el curso fue un éxito?
3. El 40% de las mujeres y el 30% de los hombres están a favor de cambios en el plan de estudios. ¿Es exacto informar que la mayoría de mujeres estuvo a favor de cambios en el plan de estudios?
4. El 51% de los encuestados está a favor de cambiar el plan de estudios. ¿Es exacto informar que más de la mitad de los encuestados estuvo a favor del cambio en el plan de estudios?
5. La encuesta fue diligenciada por 5 de los 20 instructores del curso. Los 5 informaron que estaban bien preparados para enseñar en el curso. ¿Es exacto decir que todos los instructores estaban bien preparados? ¿Es exacto decir que el 25% de los profesores estaban bien preparados?
6. El número de mujeres en la oficina política subió de 2 a 4 puestos en la legislatura de 50 personas. ¿Es exacto informar un incremento del 100%?
7. Los participantes en un programa de capacitación ganaron un 20% más que los no participantes en el programa. ¿Es exacto informar que el programa causó un 20% de incremento en el salario?

Ejercicio de aplicación 10.3: analizar resultados de un cuestionario

Llene esta encuesta y haga que por lo menos dos colegas la llenen también. Recopile las encuestas y contabilice los resultados. Trabajando solo o con los demás, resuma los resultados en forma narrativa, sacando algunas conclusiones sobre los hallazgos generales.

1. ¿En qué grado diría que tiene actualmente la capacidad analítica para hacer cada uno de los siguientes?

Destreza	*Poco o ningún grado*	*Algún grado*	*Grado moderado*	*Mucho grado*	*En grado extremo*
a. Diseñar una evaluación.					
b. Analizar datos.					
c. Desarrollar una encuesta.					
d. Dirigir un grupo de discusión.					
e. Facilitar una reunión de interesados.					
f. Escribir un informe de evaluación.					
g. Preparar un resumen oral.					

2. En este punto de este programa de capacitación, ¿estaría de acuerdo o en desacuerdo con cada una de las siguientes afirmaciones?

Afirmación	*Muy en desacuerdo*	*En desacuerdo*	*Ni de acuerdo ni en desacuerdo*	*De acuerdo*	*Muy de acuerdo*
a. El material es nuevo para mí.					
b. El material es interesante.					
c. La cantidad de tiempo dedicada a las conferencias es suficiente.					
d. La cantidad de tiempo dedicada a las discusiones en clase es suficiente.					
e. Los ejercicios son útiles.					
f. Estoy aprendiendo material que puedo utilizar en mi trabajo.					

3. Por favor, escriba los comentarios que tenga sobre el curso.

Referencias y lecturas adicionales

Babbie, E., F. Halley y J. Zaino. 2000. *Adventures in Social Research.* Thousand Oaks, CA: Pine Forge Press.

Busch, Carol, Paul S. De Maret, Teresa Flynn, Rachel Kellum, Sheri Le, Brad Meyers, Matt Saunders, Robert White y Mike Palmquist. 2005. *Content Analysis.* Writing@CSU, Colorado State University, Department of English, Fort Collins, CO. http://writing.colostate.edu/guides/research/content/.

Child and Adolescent Health Measurement Initiative (CAHMI). 2006. "Promoting Healthy Development Survey: Implementation Guidelines". Portland, OR: Oregon Health and Science University, Department of Pediatrics.

———. 2009. "Step 4: Monitor Survey Administration and Prepare for Analysis". http://www.cahmi.org.

Catterall, M. y P. Maclaran. 1997. "Focus Group Data and Qualitative Análisis Programs: Coding the Moving Picture as Well as the Snapshots". *Sociological Research Online* 2 (1). http://www.socresonline.org.uk/socresonline/2/1/6.html.

Chapman, Arthur D. 2005. *Principles and Methods of Data Cleaning: Primary Species-Occurrence Data.* Informe para la Global Biodiversity Information Facility, versión 1.0, Copenhague. http://www2.gbif.org/DataCleaning.pdf.

CIID (Centro Internacional de Investigaciones para el Desarrollo). 2008. *Qualitative Research for Tobacco Control, Module 6: Qualitative Data Analysis.* Ottawa, Canadá. http://www.idrc.ca/en/ev-106563-201-1-DO_TOPIC.html.

Cody, Ronald. s.f. "Data Cleaning 101". Robert Wood Johnson Medical School, Piscatawy, NJ. http://www.ats.ucla.edu/stat/sas/library/nesug99/ss123.pdf.

Constable, Rolly, Marla Cowell, Sarita Zornek Crawford, David Golden, Jake Hartvigsen, Kathryn Morgan, Anne Mudgett, Kris Parrish, Laura Thomas, Erika Yolanda Thompson, Rosie Turner y Mike Palmquist. 2005. *Ethnography, Observational Research y Narrative Inquiry.* Writing@CSU, Colorado State University, Department of English, Fort Collins, CO. http://writing.colostate.edu/guides/research/observe/.

Dehejia, Rajeev H. y Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimenal Causal Studies". *Review of Economics and Statistics* 84 (1): 151-61. http://www.nber.org/~rdehejia/papers/matching.pdf.

Denzin, N. y Y. Lincoln, eds. 2000. *Handbook of Qualitative Research.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Fienberg, S. E. 1980. The *Analysis of Cross-Classified Categorical Data.* 2ª ed. Cambridge, MA: MIT Press.

Firestone, W. 1987. "Meaning in Method: The Rhetoric of Quantitative and Qualitative Research". *Educational Researcher 16* (7): 16-21. http://edr.sagepub.com/cgi/content/abstract/16/7/16.

Gao, Jie, Michael Langberg y Lenard J. Schulman. 2006. *Analysis of Incomplete Data and an Intrinsic-Dimension Helly Theorem.* http://www.cs.sunysb.edu/~jgao/paper/clustering_lines.pdf.

Glass, G. y K. Hopkins. 1996. *Statistical Methods in Education and Psychology.* 3ª ed. Boston: Allyn and Bacon.

Greene, J. C. y V. J Caracelli. 1997. *Advances in Mixed Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms.* New Directions for Evaluation N° 74. San Francisco: Jossey-Bass.

Greene, J. C., V. J. Caracelli y W. F. Graham. 1989. "Toward a Conceptual Framework for Mixed-Method Evaluation Designs". *Educational Evaluation and Policy Analysis* 11 (3) 255-74.

Hawkins, Penny. 2005. "Thinking about Mixed Method Evaluation". International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio.

Jaeger, R. M. 1990. *Statistics: A Spectator Sport.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Krippendorf, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Loughborough University Department of Social Sciences. 2007. *New Methods for the Analysis of Media Content. CAQDAS: A Primer.* Leicestershire, Reino Unido. http://www.lboro.ac.uk/research/mmethods/research/software/caqdas_primer.html#what.

Miles, Mathew. B. y A. Michael Huberman. 1994. *Qualitative Data Analysis an Expanded Sourcebook.* 2ª ed. Thousand Oaks, CA: Sage Publications.

Morse, Janice M. y Lyn Richards. 2002. "The Integrity of Qualitative Research". En *Read Me First for a User's Guide to Qualitative Methods,* ed. J. M. Morse y L. Richards, 25-46. Thousand Oaks, CA: Sage Publications.

Narayan, Deepa. 1996. "Toward Participatory Research". Documento técnico del Banco Mundial 307, Washington, DC. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1996/04/01/000009265_3980625172923/Rendered/PDF/multi0page.pdf.

NCSTE (Chinese National Centre for Science and Technology Evaluation) e IOB (Policy and Operations Evaluation Department. 2006. *Country-Led Joint Evaluation of the ORET/MILIEV Programme in China.* Amsterdam: Aksant Academia Publishers.

Neuendorf, Kimberly A. 2006. *The Content Analysis Guidebook Online.* http://academic.csuohio.edu/kneuendorf/content/index.htm.

O'Rourke, Thomas W. 2000a. "Data Analysis: The Art and Science of Coding and Entering Data". *American Journal of Health Studies* 16 (3): 164-66. http://Findarticles.com/p/articles/mi_m0CTG/is_3_16/ai_72731731.

———. 2000b. "Techniques for Screening and Cleaning Data for Analysis". *American Journal of Health Studies* 16: 217-19. http://Findarticles.com/p/articles/mi_m0CTG/is_4_16/ai_83076574.

PANDA (Practical Analysis of Nutritional Data). 2000. "Chapter 2: Data Cleaning". http://www.tulane.edu/~panda2/Analysis2/datclean/dataclean.htm.

Patton, Michael. Q. 2002. *Qualitative Research and Evaluation Methods.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Porteous, Nancy. 2005. "A Blueprint for Public Health Management: The ToolKit's Approach to Program Evaluation". Presentación en el IPDET, junio-julio 2005. Ottawa, Ontario.

Porteous, Nancy L., B. J. Sheldrick y P. J. Stewart. 1997. *Program Evaluation Tool Kit: A Blueprint for Public Health Management.* Ottawa-Carleton Health Department, Ottawa. http://www.phac-aspc.gc.ca/php-psp/tookit.html.

Rahm, Erhard y Hong Hai Do. 2000. "Data Cleaning: Problems and Current Approaches". University of Leipzig, Alemania. http://homepages.inf.ed.ac.uk/wenfei/tdd/reading/cleaning.pdf.

Rossman, G. B. y B. L. Wilson. 1994. "Numbers and Words Revisited: Being 'Shamelessly Methodologically Eclectic.'" *Quality and Quantity* 28: 315-27.

Sieber, S. D. 1973. "The Integration of Fieldwork and Survey Methods". *American Journal of Sociology* 78 (6):1335-59.

Smith, J. A. 1991. *The Idea Brokers: Think Tanks and the Rise of the New Policy Elite.* Nueva York: Free Press.

StatSoft. 2008. *Electronic Textbook.* http://www.statsoft.com/textbook/glosp.html.

Stemler, Steve 2001. "An Overview of Content Analysis". *Practical Assessment, Research & Evaluation* 7 (17). http://PAREonline.net/getvn.asp?v=7&n=17.

Swimmer, Gene. 2006. "Qualitative Data Analysis, Part I". *IPDET Handbook 2005.* International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio.

University of the West of England. 2006. *Analysis of Textual Data.* Bristol. http://hsc.uwe.ac.uk/data-analysis/qualTextData.asp.

U.S. GAO (General Accounting Office). 1996. *Content Analysis: A Methodology for Structuring and Analyzing Written Material.* GAO/PEMD-10.3.1. Washington, DC. (Disponible sin costo).

Washington State University. 2000. A *Field Guide to Experimental Designs: What Is an ANOVA*? Tree Fruit Research and Extension Center. Wenatchee, WA. http://www.tfrec.wsu.edu/ANOVA/basic.html.
Weber, Robert Philip 1990. *Basic Content Analysis.* 2ª ed. Thousand Oaks, CA: Sage Publications.
Weiss, Christopher y Kristen Sosulski. 2003. *Quantitative Methods in Social Science: QMSS E-lessons.* Columbia Center for New Media Teaching and Learning, Columbia University, Nueva York. http://www.columbia.edu/ccnmtl/projects/qmss/anova_about.html.
Wolcott, H. F. 1990. "On Seeking –and Rejecting– Validity in Qualitative Research". En *Qualitative Inquiry in Education: The Continuing Debate,* eds. E. W. Eisner y A. Peshkin, 121-52. Nueva York: Teachers College Press.

Sitios web

Textos y tutoriales en línea

CAQDAS (Computer-Assisted Qualitative Data Analysis Software). 2008. http://caqdas.soc.surrey.ac.uk/.
Lane, D. M. *Hyperstat Online Textbook.* http://davidmlane.com/hyperstat/index.html.
Statistics at Square One. http://bmj.bmjjournals.com/collections/statsbk/index.shtml.
Stat Primer. http://www2.sjsu.edu/faculty/gerstman/Stat Primer.

Software informático para datos cualitativos

AEA (American Evaluation Association). http://www.eval.org/Resources/QDA.htm.
AnSWR. http://www.cdc.gov/hiv/topics/surveillance/resources/software/answr/index.htm (developer site), http://www.cdc.gov/hiv/software/answr/ver3d.htm (descarga gratis del sitio).
Atlas-ti. http://www.atlasti.com/.
CDC EZ-Text. http://www.cdc.gov/hiv/topics/surveillance/resources/software/ez-text/index.htm (developer site), http://www.cdc.gov/hiv/software/ez-text.htm (descarga de prueba gratis del sitio)
Ethnograph. http://www.qualisresearch.com/default.htm.
Friese, Susanne. 2004. *Overview/Summary of Qualitative Software Programs.* http://www.quarc.de/software_overview_table.pdf.
Hyperqual. http://home.satx.rr.com/hyperqual/.
QSR Software. NVivo 8. http://www.qsrinternational.com/.
Qualpro. http://www.qualproinc.com/.

Software informático

SPSS. http://www.spss.com.
OpenStat. Version 4. http://www.statpages.org/miller/openstat/.
Tutorial para SPSS v. 11.5. http://www.datastep.com/SPSSTraining.html/.
Iniciación en SPSS para Windows. http://www.indiana.edu/~statmath/stat/spss/win/.
WinPure. Data Cleaning Software. http://www.winpure.com.

Ejemplos de análisis de datos

Banco Mundial. http://www.worldbank.org/data.
———. *World Development Indicators 2005.* http://worldbank.org/data/wdi2005.

Carleton University. *Canadian Foreign Policy Journal.* http://www.carleton.ca/npsia/cfpj.
FMI (Fondo Monetario Internacional). http://www.imf.org/external/pubs/res/index.htm.
———. http://www.imf.org/external/np/sta/index.htm.
North-South Institute. http://www.nsi-ins.ca/ensi/research/index.html.
IISD (International Institute for Sustainable Development). http://www.iisd.org/measure/default.htm.
OCDE (Organización para la Cooperación y el Desarrollo Económicos). http://www.oecd.org/dac/.
OMS (Organización Mundial de la Salud). http://www.who.int/health-systemsperformance.
———. Statistical Information System (WHOSIS). http://www.who.int/topics/statistics/en/.
PNUD (Programa de las Naciones Unidas para el Desarrollo). *Human Development Report 2002.* http://www.undp.org/hdr2002.
UNEP (Programa de las Naciones Unidas para la Protección del Medio Ambiente). http://www.grid.unep.ch.
UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura). http://www.uis.unesco.org/en/stats/stats0.htm.
UNHCR (Alto Comisionado de las Naciones Unidas para los Refugiados). http://www.unhcr.ch/cgi-bin/texis/vtx/home.
UNSD (División de Estadísticas de las Naciones Unidas). http://unstats.un.org/unsd/databases.htm.
USAID (Agencia de los Estados Unidos para el Desarrollo). http://www.usaid.gov/educ_training/ged.html.
———. http://www.dec.org/partners/eval.cfm.
U.S. Census Bureau. http://www.census.gov/ipc/www/idbnew.html.

001El camino.indd 388 08/11/2010 03:27:57 p.m.

AFRONTAR LOS DESAFÍOS

"Cuando actuamos lo mejor que podemos, no sabemos qué milagro se perfila en nuestra vida, o en la vida de otro."

—HELEN KELLER

Capítulo 11: Evaluación de intervenciones complejas

- Visiones generales de la evaluación para el desarrollo.
- Evaluaciones conjuntas.
- Evaluaciones de programas de países.
- Evaluaciones de programas sectoriales.
- Evaluaciones temáticas.
- Evaluación de programas de alianzas mundiales y regionales.

001El camino.indd 389 08/11/2010 03:27:57 p.m.

001El camino.indd 390 08/11/2010 03:27:57 p.m.

CAPÍTULO 11

Evaluación de intervenciones complejas

Las intervenciones para el desarrollo se hacen más complejas y crean nuevos desafíos, expectativas, paradigmas para tratar con la pobreza, métodos y clientes para los evaluadores. En este capítulo se discuten las evaluaciones a nivel conjunto, de país, temáticas, sectoriales y mundiales.

Visión general de la evaluación para el desarrollo

Algunas veces, en la evaluación se debe contar con una visión general, lo cual significa ir más allá de evaluar un solo proyecto, un programa o una política para evaluar también proyectos, programas y políticas relacionados.

En la visión general se puede tratar la experiencia general y el impacto de las intervenciones para el desarrollo dentro de cualquier sector; salud, educación o de transporte. Por ejemplo, un ministerio desea establecer el impacto general y las enseñanzas aprendidas de intervenciones dirigidas a mejorar el bienestar de los niños y las mujeres; un donante desea examinar la eficacia de su estrategia para el sector educativo.

Crecientemente, como respuesta a la Declaración de París y al Acuerdo de Agra, los programas de préstamos de donantes se están basando en enfoques sectoriales (ES).

001El camino.indd 391 08/11/2010 03:27:57 p.m.

Muchos factores complejos económicos, políticos y sociales afectan las actividades y las evaluaciones para el desarrollo. El desarrollo y la evaluación para el desarrollo son cada vez más empresas pluralistas que incluyen la participación de actores no gubernamentales, como el sector privado, las agencias no lucrativas de implementación y defensa, y la sociedad civil en general.

Las intervenciones para el desarrollo no siguen una cadena lineal. Mientras existe un razonamiento del modelo "si-entonces" detrás de las intervenciones para el desarrollo, muchos factores –económicos, políticos, climáticos y sociales– interactúan con un programa y afectan sus efectos directos. Los evaluadores necesitan identificar estos factores y diseñar evaluaciones que consigan extraer su influencia.

Cuando aumenta la complejidad de la evaluación, la interacción de estos factores se hace también más compleja. Los países y sus asociados buscan determinar los efectos acumulativos de ocasionar cambios en un sector, un país, o en cuestiones temáticas multisectoriales; por ejemplo, el cambio climático.

Los evaluadores deben manejar las evaluaciones ante esta complejidad. Pero la complejidad de la evaluación no está aumentando simplemente por la ampliación del alcance de lo que se evalúa. Con mayor percepción de la carga de las evaluaciones en inexpertos o débiles capacidades en los países, la Declaración de París y el Acuerdo de Agra han reforzado la necesidad de evaluaciones de asociados múltiples de la calidad de la asistencia en un país y los efectos directos de esa asistencia.

El trabajo para el desarrollo ha evolucionado hacia una agenda más integral, atendiendo crecientemente las reformas políticas en los países, la formación de capacidad y las preocupaciones globales. A su vez, la evaluación se ha expandido,

- Reorientando su enfoque de sólo a nivel de un proyecto, un programa o una actividad hacia el nivel de país, sector temático, regional o global.
- Determinando cómo agregar mejor los efectos directos de las intervenciones a nivel de actividad y país, a fin de evaluar efectos directos globales o programáticos.
- Encontrando formas de evaluar la influencia del diseño de los programas, el enfoque de la asociación y la gobernabilidad en los resultados generales.
- Buscando la posibilidad de réplica a nivel superior y la aplicabilidad a nivel del sistema (Heath, Grasso y Jonson, 2005).

Las evaluaciones de países constituyen una forma de obtener un conocimiento general de lo que está sucediendo, y aportar percepciones sobre el efecto general y la experiencia del desarrollo dentro de un país, pero no siempre es claro cómo éstas deberían realizarse. Las evaluaciones sectoriales o temáticas –en un solo país o en varios– pueden también revelar la visión general. Es probable que utilicen varios métodos, incluyendo alguna combinación de datos disponibles, entrevistas, visitas de campo, encuestas y grupos de discusión.

En este capítulo se tratan cinco clases de visión general:

- Evaluaciones conjuntas.

- Evaluaciones de programas de países
- Evaluaciones de programas sectoriales.
- Evaluaciones temáticas.
- Evaluación de programas asociativos globales y regionales.

Evaluaciones conjuntas

Las evaluaciones conjuntas son evaluaciones en las cuales participa más de una organización.

> Existen varios grados de 'conjunto', dependiendo del punto al cual cooperen los asociados individuales en el proceso de evaluación, combinen sus recursos de evaluación y su información de evaluación. Las evaluaciones conjuntas pueden contribuir a superar problemas de atribución al evaluarse la eficacia de los programas y estrategias, la complementariedad de los esfuerzos apoyados por distintos asociados, la calidad de la coordinación en la ayuda, etcétera (OCDE, 2002, p. 26).

Las evaluaciones conjuntas se han realizado desde principios de la década de 1990, y su número, al igual que su calidad, ha aumentado desde la Declaración de París en 2005 y el Acuerdo de Agra en 2008. Sin embargo, el concepto visualizado de depender de los sistemas del país para datos basados en efectos directos es aún débil. En la evaluación de la Declaración de París indica que las evaluaciones conjuntas se están incrementando y que cada vez más son consideradas de mejor calidad y más útiles. No obstante, los donantes, sin capacidad de contar con sistemas de país robustos para los datos, todavía sienten presión para informar sobre los efectos directos bajo sus propios sistemas. Las evaluaciones conjuntas pueden ser realizadas por:

Evaluación conjunta: Evaluación en la cual participa más de una organización.

- Donante + donante.
- Donante + asociado del país.
- Varios donantes + varios asociados.
- Asociado + asociado (Chen y Slot, 2007).

Breier (2005) identifica una tipología para evaluaciones conjuntas basada en el modo de trabajo. Con un arreglo clásico de varios asociados, la participación está abierta a todos los asociados interesados, quienes participan y contribuyen activamente y en términos de igualdad. Con un arreglo de varios asociados calificado, la participación solamente está abierta para quienes cumplen ciertos requisitos, por ejemplo, pertenecer a un grupo particular (como la Unión Europea), o tener un interés fuerte en el tema que se evalúa.

Planear una evaluación conjunta es crítico para su éxito. De acuerdo con Breier:

> Una de las enseñanzas casi universales adquirida en la reciente experiencia con evaluaciones conjuntas es que es imperativo permitir tiempo suficiente al principio del proceso para desarrollar y acordar este marco de entendimiento común sobre el propósito, los objetivos, el foco y el alcance de la evaluación propuesta. Cuando no se hace esto con el tiempo y la

> paciencia necesarios, como sucedió con varias de las evaluaciones analizadas, hay una gran probabilidad de que el proceso de evaluación se vea en dificultades más adelante (2005, p. 40).

Cuando en una evaluación conjunta se incluyen solamente unas pocas agencias, la estructura de administración puede ser simple:

- Los evaluadores pueden decidir reunirse regularmente y compartir las decisiones de administración.
- Los evaluadores pueden decidir que las agencias participen por igual en la administración, pero colocar a una o más agencias en el papel de dirigencia.
- Los evaluadores pueden decidir delegar la responsabilidad de la administración a una agencia, permitiendo que las demás revisen los productos principales (OCDE, 2006).

Para evaluaciones conjuntas mayores, la estructura de administración más común es un,

> sistema de dos capas, que consiste en a) un comité de dirección de amplia membresía, y b) un grupo más pequeño de administración que opere las actividades diarias de la evaluación. Dentro de esta estructura hay libertad significativa para decidir si algunas agencias participarán como asociados silenciosos, a qué nivel de detalle el comité de dirección debe participar en la toma de decisiones, cuántos asociados deberían pertenecer al grupo de administración y cuánta responsabilidad debe delegarse (OCDE, 2006, p. 20).

Otra opción es la de un enfoque flexible o descentralizado. Cada agencia puede administrar subcomponentes discretos de la evaluación general, en una especie de estilo de rompecabezas. Es posible también adoptar un enfoque mixto, en el cual se operan algunas partes conjuntamente y otras las operan asociados específicos (OCDE, 2006).

Tanto las estructuras descentralizadas como las centralizadas tienen fortalezas y debilidades. Descentralizar la administración de una evaluación por temas facilita la delegación o la división de responsabilidades, lo que resulta en un proceso administrativo más eficiente. Sin embargo, las estructuras descentralizadas pueden crear también duplicación de esfuerzos u ocasionar que se omitan asuntos importantes. Utilizar una estructura de administración centralizada permite que cada asociado tenga aporte en todos los componentes del proceso de evaluación e influya en ellos, pero carece del sentido de propiedad y urgencia que se tiene con una estrategia descentralizada. Los asociados deben considerar los pros y los contras de cada estructura, y luego decidir cuál se ajusta mejor a la necesidad de la evaluación.

El Comité de Asistencia para el Desarrollo (CAD), identifica las siguientes áreas clave sobre las cuales deben llegar a un acuerdo los asociados en la evaluación:

- Reglas de terreno para administrar la evaluación.
- Términos de referencia para la evaluación.
- Técnicas de recopilación y análisis de datos.

- Medios de seleccionar el equipo de evaluación (oferta y contratación).
- Presupuesto, costos y financiación.
- Presentación de los hallazgos, incluyendo reglas de divulgación (OCDE, 2006).

Breier (2005) discute formas de afrontar las cuestiones legales para la evaluación conjunta. Entre estas cuestiones están las siguientes:

- Necesidades contractuales de los distintos asociados.
- Procedimientos acordados que normalmente reflejen el sistema legal, los requerimientos y las prácticas establecidas de la agencia que asuma el liderazgo en representación del grupo.
- Acuerdos de suma global frente a negociación de contratos, incluyendo cláusulas de cancelación (por desempeño deficiente).
- Estipulación de remitir informes de progreso, demostrando que los fondos se utilizan en la forma apropiada.

Una vez establecidas la estructura formal y las reglas de terreno, la dirección de la evaluación conjunta debe tornar a las cuestiones operativas y de administración. Freeman (2007) describe 16 reglas para organizar y administrar un equipo de evaluación externa (cuadro 11.1).

Otro recurso valioso es *Guidance for Managing Joint Evaluations* del CAD. (Al final del presente capítulo se suministra un vínculo al sitio web de la guía.)

Evaluaciones de programas de países

Evaluación del programa de país: Evaluación del programa completo de ayuda a un país.

Las visiones generales se centran a menudo en la ayuda a un país. Con una *evaluación del programa de país* (llamadas algunas veces evaluaciones de asistencia al país, o evaluaciones de programa de país) se evalúa el programa completo de ayuda a un país.

Una evaluación del programa de país constituye en gran parte un estudio normativo que compara lo que se hace con lo que fue planeado. Busca:

- Evaluar la pertinencia estratégica del programa de asistencia de país relativa a las necesidades del país.
- Probar la implementación de los efectos directos de toda la agencia para determinar si se obtuvieron los efectos directos previstos.
- Identificar los éxitos y fracasos en diferentes sectores o de diferentes enfoques utilizados en el país, e identificar los factores que contribuyeron al desempeño.
- Identificar la eficacia de la ayuda del donante a un país dado (OCDE, 1999).

Las evaluaciones de asistencia a un país se centran usualmente en los criterios del CAD de pertinencia, eficiencia, impacto y sostenibilidad, y tienen en cuenta el desempeño del donante, el del país o ambos (recuadro 11.1).

Cuadro 11.1 *Seleccionar, organizar y administrar un equipo de evaluación externa e informar los efectos directos*

Asunto	*Regla*
Seleccionar el equipo de evaluación externa	• Asegurar que la pericia central en evaluaciones complejas de gran escala de la cooperación para el desarrollo sea lo más sólida posible. • Mantener la organización tan sencilla como sea posible, y trabajar con organizaciones con las que se haya trabajado antes, también en cuanto sea posible. • Integrar consultores nacionales al proceso de licitación competitiva internacional e incluirlos en la selección y el diseño de la metodología. • En estudios de varios países, hacer que cada equipo de campo combine recursos de distintas organizaciones en los consorcios, y no que cada organización se especialice geográfica o institucionalmente.
Administrar el equipo de evaluación externa	• Hacer claro el compromiso con la evaluación a nivel de junta directiva de la principal organización de la evaluación. • Mantener talleres del equipo de evaluación, a fin de desarrollar un enfoque común para la medición y la presentación de la información. • En cuanto sea posible, crear un grupo de administración que informe al comité de dirección de la evaluación general. • Permitir tiempo suficiente en las reuniones del comité de dirección para discusiones completas, y adoptar una posición común. • En evaluaciones conjuntas, asegurar que los evaluadores externos operen abiertamente y con transparencia, y vinculen directamente los métodos con la evidencia acumulada para los hallazgos, conclusiones y recomendaciones. • En las negociaciones para obtener recursos adicionales, cuando sean necesarios, hacer que el equipo de evaluación y el grupo de administración empiecen con un acuerdo sobre la división del trabajo que deba emprenderse como resultado del contrato original (y utilizar el conjunto de recursos originales), y el trabajo que sea resultado de nuevos asuntos e intereses, o que surja de circunstancias imprevisibles. Esto requerirá que el equipo prepare planes detallados con costos y tiempos para cualquier nuevo trabajo requerido.
Organizar el equipo de evaluación externa	• Asegurar que la organización que lidera la evaluación en el consorcio tenga un compromiso fuerte y un historial en evaluación de cooperación para el desarrollo internacional. El proyecto debe ser una adecuación natural a su negocio y mercado centrales. • Asegurar que las partes interesadas mantengan una orientación positiva a través de toda la evaluación. • Asegurar que el equipo de evaluación externa que trata con todos los integrantes del comité de dirección evite demostrar sesgos institucionales y se comporte como si todos fuesen iguales. • Invertir una proporción sustancial del presupuesto en actividades de difusión y seguimiento.
Informar los efectos directos del equipo de evaluación externa	• Presentar hallazgos preliminares antes del borrador del informe. • Hacer que los evaluadores y los integrantes del comité de dirección discutan los borradores con actitud abierta hacia mejoras que puedan hacerse, asegurando que los evaluadores puedan mantener su responsabilidad objetiva para con los hallazgos y conclusiones de la evaluación y estén dispuestos para eso.

Fuente: basado en Freeman, 2007.

En las evaluaciones de programas de países se pueden encontrar problemas sustanciales, como los siguientes:

- La ayuda general a un país carece de objetivos coherentes y efectos directos bien articulados, y refleja en cambio un enfoque oportunista.
- Varias fuentes financian intervenciones para el desarrollo similares, lo que dificulta la atribución. (Nota: la cofinanciación es una cosa, pero la financiación de intervenciones similares es otra.)
- Usualmente no existe cartografía en la asistencia dentro del país, lo que dificulta conocer lo que otros están haciendo en el área de la intervención.
- Al igual que en cualquier evaluación, "las reputaciones están en juego y los temores de las consecuencias pueden amenazar el entusiasmo y el compromiso del personal del programa y los asociados. La evaluación de programas de países, como cualquier evaluación, debe proceder con sensibilidad" (OCDE, 1999, p. 18).

La CAD Network on Development Evaluations propone las siguientes recomendaciones:

- Una mayor proporción de las evaluaciones debe realizarse conjuntamente, con participación completa y activa de los receptores de la ayuda y los demás asociados.
- Los países en desarrollo deben mostrar mayor iniciativa para asumir la dirección en la planeación, la coordinación y la programación de las intervenciones.
- Debe apoyarse a los países en desarrollo en la formación de su capacidad institucional para iniciar y liderar evaluaciones conjuntas.
- Se requiere compartir de mejor manera la coordinación y el conocimiento entre los distintos asociados dentro de los países receptores de la ayuda. Deben construirse y expandirse las redes nacionales de seguimiento y evaluación y las asociaciones profesionales.
- Al emprender una gran evaluación conjunta con la participación de varios países en desarrollo, debe facilitarse a esos países una reunión para coordinar sus puntos de vista y sus aportes (OCDE, 2005, p. 7).

El Grupo de Cooperación para la Evaluación de los Bancos de Desarrollo Multilateral publicó hace poco tiempo estándares de buenas prácticas para las evaluaciones de países. Con estos estándares se aspira armonizar más los criterios, procesos y procedimientos de evaluación, y hacer que los efectos directos de las evaluaciones sean más comparables. Al final del presente capítulo se provee un vínculo a un sitio web para la versión 2003 de estos estándares (Evaluation Cooperation Group, 2008).

La evaluación de programas de país debería iniciarse con términos de referencia claramente definidos, desarrollados en conjunto por el país y la organización para el desarrollo, a fin de determinar exactamente qué esperan las partes interesadas. Debe especificarse también el tiempo que cubrirá y los términos de referencia deberían:

Recuadro 11.1 *Ejemplo de metodología para evaluación del programa de país*

Desde 1995, el Grupo de Evaluación Independiente (IEG) del Banco Mundial ha emprendido más de 70 evaluaciones de programas de países, y ha desarrollado una metodología claramente articulada (Banco Mundial, 2008). La metodología de evaluación del IEG es un enfoque tanto ascendente como descendente. Para cada uno de los objetivos principales, la Country Assistance Evaluation evalúa:

- La pertinencia del objetivo.
- La pertinencia de la estrategia del Banco para el cumplimiento del objetivo, incluyendo el equilibrio entre instrumentos de préstamo y otros instrumentos.
- La eficacia con que fue implementada la estrategia.
- Los efectos directos obtenidos.

La evaluación se hace en dos pasos. El primero es una revisión descendente sobre si el programa del Banco cumplió un objetivo particular o un resultado planeado de éste, y causó un impacto sustancial en el desarrollo del país. El segundo es una revisión ascendente de los productos y servicios del Banco (préstamos, servicios de análisis y asesoría, y coordinación de la ayuda) utilizados para lograr los objetivos. En conjunto, con estos dos pasos se verifica la consistencia de los hallazgos de los productos y servicios y las dimensiones del impacto en el desarrollo. Posteriormente se realiza una evaluación de la contribución relativa a los efectos directos obtenidos por el Banco, otros donantes, el gobierno y factores exógenos.

Cuando el IEG evalúa el impacto esperado en el desarrollo de un programa de ayuda, calibra el grado al cual fueron pertinentes y se lograron los principales objetivos estratégicos. En los programas se expresan usualmente sus objetivos, utilizando los de alto orden, por ejemplo, los Objetivos de Desarrollo del Milenio (ODM) y la reducción de la pobreza. La Country Assistance Strategy (Estrategia de Asistencia de País) puede fijar también metas intermedias, como una mejor focalización de los servicios sociales o la promoción del desarrollo rural integrado. Puede especificar también cómo se espera que los programas contribuyan a lograr el objetivo de alto orden.

Con la evaluación se busca validar si los objetivos intermedios produjeron beneficios netos satisfactorios, y si la cadena de resultados especificada en la Country Assistance Strategy fue válida. Si no se especificaron vínculos causales, el evaluador debe intentar reconstruir esta cadena causal a partir de la evidencia. El evaluador debe también evaluar la pertinencia, la eficacia y el resultado de los objetivos intermedios y de alto orden.

Los evaluadores evalúan también el grado al cual los clientes demuestran propiedad de las prioridades del desarrollo internacional. Ejemplos de estas prioridades incluyen los ODM, las salvaguardias para los derechos humanos, la democratización, la igualdad de géneros y el medio ambiente. Idealmente, la Country Assistance Strategy identifica y atiende estos asuntos, permitiendo al evaluador concentrarse en si los enfoques adaptados para esta asistencia fueron apropiados. La estrategia puede haber restado importancia a ciertos conflictos, o haber evitado atender restricciones clave del desarrollo del cliente, reduciendo de manera potencial la aplicabilidad del programa y la propiedad del cliente, y aumentando los efectos colaterales no previstos, por ejemplo, violaciones a las salvaguardias. Todos estos factores deben tenerse en cuenta al juzgar los efectos directos del programa. El punto importante es que aun si los proyectos del Banco Mundial tuvieron un buen desempeño, si el país no lo tuvo, la asistencia puede juzgarse como insatisfactoria, o viceversa.

001El camino.indd 398 08/11/2010 03:27:57 p.m.

- Expresar con claridad el propósito de la evaluación, sus criterios y la forma como se utilizarán los hallazgos.
- Especificar las prioridades originales de la organización para el programa de país (por ejemplo, reducción de la pobreza, mayor producción de cultivos).
- Especificar la presentación de información, su difusión y los procedimientos de seguimiento; lo ideal es una divulgación completa de los hallazgos.

Debido a la dificultad de encontrar un caso hipotético alternativo, es importante efectuar pruebas de comparación en las evaluaciones de programas de países. Las comparaciones se hacen usualmente con países similares, por lo general de la misma región. A medida que más evaluaciones de programas se divulgan públicamente, serán más factibles más comparaciones entre organizaciones. Como se señaló en el GPS, la complejidad de las evaluaciones de países está en crecimiento. Por ejemplo, como repuesta a la Declaración de París y el Acuerdo de Agra, existe la necesidad de evaluaciones de varios asociados de la ayuda a los países, de modo que se extienda más allá de los Banco Multilateral de Desarrollo (BMD) e incluya todas las fuentes de ayuda externa, para las cuales los problemas de la evaluación son significativamente mayores.

Evaluaciones de programas sectoriales

Evaluación de programas sectoriales: Evaluación que se realiza sobre un sector importante de programas, por ejemplo, educación, salud, vivienda o transporte.

Son evaluaciones que se realizan sobre un sector importante de programas, por ejemplo, educación, salud, vivienda o transporte. Los Lineamientos de Evaluación de la Organización Internacional para las Migraciones (OIM), definen la evaluación de un sector como "una evaluación de una variedad de acciones de ayuda, todas las cuales se encuentran ubicadas en el mismo sector, ya sea en un solo país o en varios. Un sector cubre un campo específico de actividades, por ejemplo, salud, industria, educación, transporte o agricultura" (OIM, 2006, p. 30).

Dado que las evaluaciones de programas sectoriales comprenden muchos proyectos con distintos objetivos y diferentes donantes, son más complejas que las evaluaciones de programas de países (y aún más si incluyen varios países). A semejanza de estas últimas, las evaluaciones sectoriales son generalmente normativas.

Es útil comparar evaluaciones de programas de países con evaluaciones de programas sectoriales en el IEG. A nivel de países, el IEG realiza usualmente una revisión del portafolio completo de proyectos del Grupo del Banco Mundial en el período de tiempo aplicable al país. En las evaluaciones sectoriales, el IEG realiza una revisión del portafolio de todos los proyectos en el sector en el período de tiempo aplicable a los países seleccionados. Sin embargo, en el caso de revisiones sectoriales, el IEG se encuentra a menudo con problemas de definición. Si, por ejemplo, un proyecto cubre varios sectores, ¿el IEG incluye solamente los proyectos para los cuales el sector de interés fue el sector principal, o incluye también aquellos para los cuales fue un tema secundario o

terciario? Si el IEG toma el conjunto más amplio, ¿pondera los efectos directos según el porcentaje de los recursos que van a ese sector frente a los demás? Así, aunque tanto las evaluaciones de países como las sectoriales incluyen revisiones de portafolio, los temas pueden ser más complejos para las revisiones sectoriales.

Problemas similares pueden suscitarse sobre el diseño y los métodos de la evaluación. Para el Grupo del Banco Mundial, y de manera creciente para otros prestamistas y donantes, los métodos de evaluaciones de países se encuentran ya más estandarizados. En contraste, las evaluaciones sectoriales tienden a ser más *sui géneris*, determinándose el diseño y los métodos por las preguntas de evaluación específicas y los recursos disponibles para el estudio. Aunque la mayoría incluye casos de estudio de países, esto no constituye una necesidad. Cuando se utilizan casos de estudio, existen cuestiones importantes de selección de casos y métodos que deben resolverse.

La propiedad y la responsabilidad de los asociados son asuntos clave en las evaluaciones de programas sectoriales (Danish Ministry of Foreign Affaires, 1999). Tanto las organizaciones para el desarrollo como la institución asociada comparten la preocupación por mejorar la prestación de la ayuda, la responsabilidad y la sostenibilidad en el sector. La breve descripción de la evaluación conjunta del sector de la salud de Tanzania resalta estas preocupaciones básicas (recuadro 11.2).

Evaluaciones temáticas

Tratan con "aspectos o temas seleccionados en varias actividades para el desarrollo" (Danish Ministry of Foreign Affaires, 1999, p. 30). Estos temas surgen de las declaraciones de políticas. La organización para el desarrollo puede decidir, por ejemplo, que todos los proyectos o los programas atiendan uno o varios problemas específicos: de género, ambientales, de sostenibilidad social o de alivio de la pobreza. Cada uno de estos problemas debe abordarse en todas las etapas del proyecto o del programa y para todas las formas de ayuda.

Evaluación temática: Evaluación de aspectos o de temas seleccionados en varias actividades para el desarrollo.

Al igual que con las evaluaciones de países y sectoriales, el enfoque en las evaluaciones temáticas es ascendente y descendente. Se evalúan los temas proyecto por proyecto; la información de estas evaluaciones ofrece una riqueza de información para las evaluaciones temáticas. Además, en las evaluaciones temáticas se va más allá del nivel de proyectos y se examinan todos los proyectos; por lo general, también en ellas se seleccionan los países para estudios más profundos (casos de estudio). En una evaluación temática se examinan muchas clases distintas de información (recuadro 11.3) y luego se extrae la información acumulada de estas fuentes.

Uno de los hallazgos básicos de la evaluación temática sobre niños que buscan en la basura fue el papel del género: la mayor parte de los niños que lo hacían eran niñas y casi todos los adultos eran mujeres. Este estudio se diferenció por la sensibilidad que mostró sobre los problemas de género.

001El camino.indd 400 08/11/2010 03:27:57 p.m.

Recuadro 11.2 *Evaluación conjunta externa del sector de salud de Tanzania, 1999-2006*

Durante los años noventa, el sector de la salud de Tanzania sufrió un período de estancamiento. Los servicios locales de salud se caracterizaron por una severa escasez de medicamentos esenciales, equipos y suministros, al igual que de deterioro en la infraestructura, una administración deficiente, falta de supervisión y de motivación para el personal. El sector sufrió el estancamiento o el deterioro de los servicios hospitalarios. En la prestación de los servicios de salud existió poca cooperación entre el sector público, las organizaciones religiosas y los proveedores privados de servicios. La financiación del sector fue sumamente escasa y el sector público solamente gastaba la suma de 3,46 dólares per cápita. Hubo también poca evidencia de coordinación de apoyo al sector de la salud por parte de los asociados para el desarrollo.

El gobierno de Tanzania y sus asociados para el desarrollo (Bélgica, Canadá, Dinamarca, Alemania, Países Bajos y Suiza) respondieron a esta situación en un proceso que se inició con una misión de planeación conjunta acordada por el gobierno hacia la mitad de la década. Para 1999, este proceso había producido el primer gran plan estratégico del sector de la salud, el Programa de Trabajo para el Sector de la Salud (POW), y un acuerdo para que el apoyo al sector de la salud tuviera lugar dentro del marco de un enfoque sectorial (ES). El POW y el posterior Health Sector Strategic Plan 2 (HSSP2) articularon un proceso de reforma al sector de la salud dirigido a resolver las deficiencias evidentes en el sector y a lograr metas y objetivos específicos en la salud, según se especifican en los ODM y en la Estrategia Nacional para el Crecimiento y la Reducción de la Pobreza (NSGRP/Mkukuta).

El POW y el HSSP2 identificaron varias áreas prioritarias de intervención estratégica:

- Fortalecer los servicios de salud en los distritos y reformar y fortalecer los hospitales regionales y nacionales de remisiones.
- Transformar el papel del Ministerio de Salud y Bienestar Social central en una organización política de facilitación.
- Mejorar los sistemas centrales de apoyo (incluyendo infraestructura, sistemas de información para la administración de la salud, suministro de medicamentos, transporte, comunicaciones y tecnología de información).
- Adoptar e implementar una estrategia nacional para combatir el VIH/Sida, como elemento explícito del HSSP2.
- Mejorar las relaciones entre el gobierno y los asociados para el desarrollo, a fin de enriquecer la armonización y alineación de los recursos externos y nacionales en una asociación más eficaz.

La evaluación se concentró en cuatro áreas:

- La aplicabilidad de los planes estratégicos y de implementación del sector de la salud para lograr los ODM en salud y para las metas y objetivos NSGRP del sector de la salud y la adecuación y la pertinencia del apoyo externo.
- El grado de progreso y de logros en cada una de las nueve prioridades estratégicas del proceso de reforma del sector de la salud.

(Continúa en la página siguiente)

(Continúa en la página siguiente)

- Los logros en el mejoramiento del acceso, la calidad del servicio y los efectos directos en salud durante el período de evaluación.
- Los cambios en la asociación durante el período de evaluación, incluyendo la evolución de los esfuerzos para la armonización, alineación y el uso de diferentes modalidades de ayuda.

Entre diciembre de 2006 y septiembre de 2007, un equipo de ocho consultores internacionales en salud y evaluación, tres de ellos nacionales de Uganda y Malawi, realizó la evaluación. Las principales metodologías utilizadas incluyeron las siguientes:

- Una extensa revisión de documentos.
- Entrevistas a informantes clave de nivel nacional.
- Autoevaluaciones de distrito efectuadas por el personal del Ministerio de Salud y Bienestar Social en 16 distritos de Tanzania.
- Casos de estudio en profundidad en seis distritos (incluyendo discusiones con integrantes de la comunidad) realizados por el equipo de la evaluación.
- Análisis de flujos financieros y otros recursos hacia el sector de la salud a nivel nacional, regional y de concejos.
- Una revisión de los datos de los efectos directos sobre salud.

La información de la evaluación de todas las metodologías utilizadas fue sometida a prueba y triangulada para identificar similitudes y fortalecer la validez de los hallazgos y conclusiones.

Fuente: Danish Ministry of Foreign Affaire, 2007.

Las características distintivas de un enfoque sensible al género son:

- Un marco conceptual en el que se reconoce la naturaleza de los géneros en el desarrollo y la contribución de la igualdad de los géneros al desarrollo económico y social.
- La creación de una base de datos sobre géneros a nivel nacional, sectorial o local que sintetice datos pertinentes a los géneros, e identifique los problemas clave de género que deben abordarse en el diseño y la evaluación temática de los proyectos.

La falta de desagregación de datos por género se utiliza con mucha frecuencia para justificar la falta de atención a los asuntos de género. Cuando los datos no se encuentran desagregados por géneros, es importante definir estrategias para desarrollar las bases de datos apropiadas y posibilitar que se realice un mejor análisis de géneros en el futuro. Tales estrategias incluyen:

- Asegurar que los métodos de recopilación de datos generen información sobre hombres y mujeres, y que los asuntos sobre géneros (como la división de géneros en el trabajo, análisis de uso del tiempo, control de los recursos y toma de decisiones a nivel de familia y de comunidad) se incorporen al diseño de la investigación.

Recuadro 11.3 *Ejemplo de evaluación temática: evaluación sobre niños que buscan en la basura en África, Asia y Europa*

La organización Waste es una entidad que trabaja para el mejoramiento del desarrollo sostenible de las condiciones de vida de los pobres urbanos, y el entorno urbano en general. La Organización Internacional del Trabajo (OIT) la contrató para hacer una evaluación temática del trabajo de niños que escarban en las basuras. El propósito de la evaluación fue ofrecer orientación a la OIT sobre el mejor tratamiento que podría darse a la explotación de los niños en este sector.

En la evaluación temática se identificaron y evaluaron críticamente la búsqueda en las basuras y varios enfoques para tratar el problema del trabajo infantil en relación con la búsqueda en las basuras. La información se tomó de varios proyectos realizados en este sector por el International Programme on the Elimination of Child Labour y de esfuerzos similares de otras agencias, instituciones o gobiernos. El trabajo dio como resultado una evaluación estratégica y recomendaciones para la OIT. (La evaluación, que lleva el nombre de *Addressing the Exploitation of Children in Scavenging (Waste Picking): A Thematic Action Agenda on Chile Labour*, puede descargarse del sitio web de la OIT.)

Los investigadores de campo de Waste, en Egipto, India, Rumania, Tanzania y Tailandia, escribieron informes adicionales basados en sus hallazgos. Por ejemplo, en Tanzania el informe incluyó una revisión de documentos pertinentes, entrevistas con los administradores y personal del proyecto, entrevistas con otras partes interesadas, una visita de campo a dos puntos de transferencia en la ciudad, y el sitio principal de descarga en Dar es Salaam, y entrevistas a escarbadores de basura y representantes de una asociación de escarbadores de basura.

Fuente: Duursma, 2004; Waste, 2005.

- Asegurar que se recopila información sobre, y proveniente de, los distintos integrantes de la familia y que el "jefe de familia" (usualmente definido como masculino) no sea la única fuente de información.
- Complementar los métodos convencionales de recopilación de datos con métodos incluyentes de género donde sea apropiado.
- Asegurar que el equipo de investigación incluya un equilibrio entre hombres y mujeres.
- Asegurar que se consulta a las partes interesadas durante el diseño, el análisis y la divulgación de la evaluación, y que las consultas incluyan grupos que representen tanto a hombres como a mujeres (Banco Mundial, 2007e).

En la planeación y la evaluación para el desarrollo se utilizan con frecuencia enfoques "neutrales con respecto al género", en los que se asume que hombres y mujeres tienen las mismas necesidades para el desarrollo. En la mayor parte de las sociedades, los hombres suelen predominar en la toma de decisiones comunitaria y familiar, de manera que los enfoques "neutrales con respecto al género" responden mayormente a las prioridades masculinas. Ignorar las necesidades y capacidades de la mujer reduce significativamente

la eficiencia y la equidad de las políticas y programas. La forma como se aplican muchas herramientas de planeación y evaluación para el desarrollo no es sensible al género, de modo que aun cuando las consultas comunitarias o las encuestas familiares tienen el propósito de capturar la opinión de todos los sectores de la comunidad, las preocupaciones de la mujer a menudo no se llegan a captar completamente.

Evaluación de programas de alianzas mundiales y regionales (GRPP)

Los GRPP constituyen una modalidad cada vez más importante para canalizar y distribuir la ayuda para el desarrollo, a fin de abordar problemas apremiantes globales o regionales. La mayor parte de los GRPP son específicos de un sector o un tema, por ejemplo, agricultura, medio ambiente, salud, finanzas o comercio internacional.

Los GRPP son asociaciones programáticas en las cuales:

- Los asociados contribuyen y agrupan recursos (financieros, técnicos, de personal y reputación) con el fin de lograr los objetivos acordados a través del tiempo.
- Las actividades del programa son de alcance global, regional o de varios países.
- Los asociados establecen una nueva organización con una estructura de gobernabilidad y unidad de administración para ejecutar estas actividades (Banco Mundial, 2007).

Usualmente se establece un nuevo GRPP después de que un grupo de actores globales o regionales identifica un problema o una oportunidad que requiere la atención pública, pero que atraviesa las fronteras soberanas nacionales. Los actores pueden ser donantes o receptores y provenir del sector público o del privado.

Considérese el ejemplo de la administración de los recursos de agua en las cuencas de los ríos Mekong o Nilo. Los problemas o las oportunidades no pueden abordarse en forma eficaz únicamente a nivel de un solo país, pues muchos países (6 en el caso del Mekong y 19 en el caso del Nilo) están involucrados en el manejo de la cuenca. Además, se producen economías sustanciales por la acción colectiva a nivel global o regional (generación de nuevas tecnologías o buenas prácticas aplicables al desarrollo en muchos países).

Los GRPP se concentran a menudo en la producción o preservación de bienes públicos globales (por ejemplo, investigación que produce una mejor tecnología de semillas o incrementa la biodiversidad), o reducción de problemas públicos globales (por ejemplo, VIH/Sida o contaminación). Los bienes públicos globales son bienes no rivales (su provisión a una persona o a un país no disminuye la oferta para otro) y no excluibles (una vez provistos a una persona o a un país, todos se benefician), y sus beneficios se extienden más allá de las fronteras nacionales.

Programa de alianza mundial y regional (GRPP): Asociación programática en la cual los asociados contribuyen y agrupan recursos para lograr objetivos acordados de alcance global, regional o de varios países.

Los asociados de los GRPP tienen usualmente perspectivas e incentivos distintos. Se configura una estructura formal de gobernabilidad que dispone objetivos formales, criterios sobre membresía o participación, procedimientos de toma de decisiones y algunas veces criterios de asignación.

Aproximadamente unos 150 GRPP se han establecido, y el nivel de desembolso oscila entre unos pocos millones de dólares y más de 1.000 millones por año. Sus actividades pueden limitarse a compartir y comunicar el conocimiento a través de redes, o pueden también abarcar asistencia técnica, inversión y actividad de mercado o comercial. Ejemplos de GRPP incluyen el Consultative Group for International Agricultural Research, el West Africa HIV/AIDS and Transport Program, la Global Water Partnership, el Integrated Framework for Trade y la Medicines for Malaria Venture.

Las características de los GRPP que hacen compleja su evaluación incluyen las siguientes:

- La naturaleza de los programas como asociaciones tiene dos implicaciones. Primero, puede ser necesario tener en cuenta las diferentes perspectivas de los distintos asociados. Segundo, la eficacia de la estructura de gobernabilidad debe evaluarse, tanto con respecto a criterios intrínsecos relativos a las buenas prácticas en las asociaciones, como en cuanto a su contribución a los efectos directos de los programas (positivos o negativos).
- A diferencia de los proyectos, los GRPP no disponen de un tiempo fijo para lograr sus objetivos. En realidad, los objetivos y estrategias evolucionan a menudo cuando aumenta (o disminuye) la financiación, cambia la composición de donantes o de la asociación, o cambian las condiciones externas.
- En los distintos niveles (local, nacional y global) se incurre en costos y se obtienen beneficios. En una evaluación de la eficacia deben tenerse en cuenta las contribuciones y la interdependencia entre los diferentes niveles, lo que hace que sea compleja la construcción de la cadena de resultados, y plantea un problema de agregación de efectos directos.
- La buena dimensión global y regional de los GRPP significa que existe una divergencia entre los costos y beneficios captados en los niveles nacional y global. Esto hace compleja la evaluación de la efectividad de costos, y también hace que sea esencial la eficacia de la coordinación global-país y los vínculos.
- Debido a que el marco de tiempo es abierto y más largo, con mucha frecuencia se solicita a los evaluadores analizar el progreso hacia los objetivos estratégicos complejos a mediano plazo, por ejemplo, la movilización de los recursos dirigida a un aumento gradual de las operaciones, la devolución a los ejecutores locales, la independencia de la agencia anfitriona, la clausura o la salida de la intervención, o patrones modificados de colaboración con nuevos programas globales emergentes.

Dada la creciente participación del Grupo del Banco Mundial en los GRPP, y su importancia cada vez mayor en el logro de los efectos directos para el desarrollo sostenible

global, en 2006 el Grupo de Evaluación Independiente (IEG) del Banco Mundial inició dos nuevas actividades a fin de apoyar el examen continuado y el mejoramiento de su desempeño. Primero, como respuesta a una solicitud de la red de evaluación OCDE/CAD (apoyada por el Grupo de Evaluación de las Naciones Unidas [UNEG] y el Grupo de Cooperación para la Evaluación de los Bancos Multilaterales de Desarrollo), el IEG orientó sus esfuerzos hacia desarrollar principios de consenso y estándares para evaluar los GRPP. El primer producto de este trabajo fue el *Sourcebook for Evaluating Global and Regional Partnership Programs* (Banco Mundial, 2007f). Hay trabajo en curso sobre investigación, que contribuirá a la publicación de un informe acompañante, presentando lineamientos más detallados y ejemplos.

Segundo, el IEG empezó a revisar evaluaciones independientes de los GRPP en las cuales participa el Banco Mundial, para evaluar la calidad de la evaluación y suministrar retroalimentación, realizar una evaluación independiente de los hallazgos y recomendaciones y evaluar el desempeño del Banco como asociado. Siete de estas revisiones de programas globales se han divulgado al público y están surgiendo enseñanzas de ellas.

La evaluación y la revisión del programa global de la Medicines for Malaria Venture (MMV) es un ejemplo de esas revisiones. Fundada en 1999, la MMV busca reducir la carga de la malaria en países de enfermedades endémicas, mediante el descubrimiento, el desarrollo y la producción de nuevas medicinas contra la malaria, económicas por intermedio de asociaciones público-privadas eficaces. Las contribuciones acumuladas fueron de 151 millones de dólares hasta finales de 2006. La Secretaría del programa, que está organizado como una organización independiente sin fines lucrativos según la ley suiza, está ubicada en Ginebra y existe una oficina satélite en Nueva Delhi. La junta de directores de la MMV integrada por 11 miembros distinguidos de la industria, la academia, la Organización Mundial de la Salud (OMS) y la Bill and Melinda Gates Foundation, que aporta el 60% de la financiación.

Los donantes de la MMV contrataron una evaluación externa en 2005. La junta cooperó con la evaluación, pero no la comisionó. La evaluación fue realizada por un equipo de cuatro personas, bajo la dirección de un distinguido profesor africano de salud pública.

Los términos de referencia exigían el uso de criterios estándar de evaluación con preguntas de evaluación, explorando más detalles sobre aspectos específicos de los GRPP en general, y la naturaleza de las actividades financiadas por la MMV en particular. La discusión siguiente ilustra cómo se aplican esos criterios en evaluaciones complejas de GRPP.

*Pertinencia***.** Se plantearon preguntas sobre la pertinencia de los objetivos del programa: ¿Hubo consenso internacional sobre el problema que se deseaba resolver y la necesidad de que el programa lo abordara? En la evaluación se señaló la pertinencia del programa para lograr el ODM de prevenir la difusión de enfermedades transmisibles, lo que es también un bien público global. En la evaluación de la pertinencia se aplicó igualmente el criterio del principio subsidiario, que expresa que los programas deben

001El camino.indd 406 08/11/2010 03:27:57 p.m.

implementarse al nivel más local posible. En la evaluación se hizo la pregunta "¿El programa presta servicios que no pueden prestarse a nivel de país o local?" Debido al bien público que se abordaba, en la evaluación se concluyó que el programa cumplía el criterio del principio subsidiario.

En la evaluación se analizó el valor agregado de MMV relativo a otras fuentes de suministro y servicios (analizando otros programas sobre la malaria y programas de desarrollo de medicamentos solamente del sector privado). Se evaluó también la pertinencia de los objetivos del programa a las necesidades de los beneficiarios. Se juzgó que el programa era aplicable según este criterio, porque atendía no solamente el desarrollo de medicamentos sino también asuntos de prestación y acceso.

En la evaluación se hicieron también las preguntas siguientes: ¿Son pertinentes la estrategia y el diseño del programa? ¿Fue pertinente y apropiada la elección de las intervenciones? ¿Cuáles fueron el alcance y los destinatarios geográficos? ¿Cómo se utilizaron las asociaciones público-privadas? ¿Habría sido más pertinente un diseño alternativo para orientar la estrategia del programa?

Eficacia (logro de los objetivos). En la evaluación se analizaron los productos y efectos directos alcanzados, en relación con las expectativas, al igual que los factores que contribuyeron o que obstaculizaron el logro de los objetivos. Este aspecto de la evaluación fue muy dependiente del marco de la calidad del seguimiento y evaluación (Sye) y la recopilación de datos. Se evaluaron las contribuciones de los factores siguientes:

- Participación de las partes interesadas.
- Eficacia de la administración del portafolio.
- Eficacia de la administración de la asociación público-privada.
- Uso de un enfoque científico.
- Eficacia de los enfoques básicos sobre el producto.
- Efectos directos imprevistos.

Eficiencia. Aunque no se les solicitó explícitamente como un criterio separado en los términos de referencia, los evaluadores reconocieron la importancia de analizar los costos de oportunidad de medios más eficientes de prestación de servicios. Analizaron los costos de manejo y administración relativos a los gastos del programa y a las tendencias en el tiempo, y los contrastaron con estándares de comparación, por ejemplo, el costo de desarrollo de medicamentos para otras enfermedades, y los costos de un proceso característico del desarrollo de un producto en una gran empresa farmacéutica privada.

Gobernabilidad. En la evaluación se analizaron la legitimidad de la gobernabilidad, examinando la representación de países endémicos sobre el cuerpo de gobernabilidad; el uso de pericia científica, como la de los investigadores de la OMS y los países endémicos; la evidencia de la confianza de los asociados en campañas exitosas de movilización de recursos y reacciones a las evaluaciones. En la evaluación se analizó también la eficacia de la gobernabilidad en el desempeño de las siguientes funciones importantes:

001El camino.indd 407 08/11/2010 03:27:57 p.m.

- Fijar la estrategia. (¿Qué tan bien informó la estructura de gobernabilidad las deliberaciones sobre la nueva dirección de atender el acceso a los medicamentos a nivel del país?)
- Colaboración con otras agencias con mandatos similares (en la evaluación se recomendó más atención a este aspecto).
- Movilización de recursos (en la evaluación se recomendó actuar más proactivamente para aumentar los recursos necesarios y diversificar el apoyo de los donantes).

Cuestiones de sostenibilidad o de estrategias a mediano plazo. En la evaluación se analizó si era probable que los beneficios del programa se sostuvieran en las condiciones actuales, atendiendo la sostenibilidad en términos de las necesidades de financiación y la defensa sostenida y apoyo de los objetivos. También se evaluó si era posible hacer cambios para que las condiciones fuesen más favorables para la sostenibilidad. (¿Debía cambiarse la escala del programa? ¿Podrían hacerse más eficaces las relaciones con otras organizaciones? ¿Debían analizarse arreglos organizativos o de gobernabilidad alternativos?) En la evaluación se trató también el mejoramiento del SyE. Los términos de referencia exigían a los evaluadores ensayar una plantilla propuesta por los donantes para evaluar la eficacia de las asociaciones público-privadas. Además, los evaluadores hicieron recomendaciones sobre el mejoramiento del seguimiento para apoyar iniciativas futuras de seguimiento y evaluación. En la metodología de la evaluación se incluyó una revisión de documentos sobre el programa, entrevistas extensas con partes interesadas a nivel de programa y de actividad, observación de una reunión del comité científico asesor y visitas al sitio en el país.

La Global Program Review evaluó la independencia y la calidad de la evaluación. Tuvo en cuenta quién comisionó la evaluación y cómo fue administrada, con el fin de juzgar el grado de independencia. Examinó si la evaluación cubrió todos los aspectos del trabajo de acuerdo con los términos de referencia, y también la calidad de la evaluación y su impacto final (el grado al cual la administración del programa puso en práctica las recomendaciones). La revisión proporcionó una evaluación independiente de los hallazgos de las evaluaciones y confirmó que proporcionaron una base sana para acciones de seguimiento de la administración del programa y el Banco.

La Global Program Review evaluó el desempeño del Banco Mundial como asociado, analizando el uso de su poder de convocatoria para establecer el programa; su apoyo financiero posterior, que fue inicialmente importante, disminuyó con el tiempo en relación con otros donantes, y la participación del Banco en las deliberaciones de los organismos de gobierno y los comités. Identificó problemas concernientes al papel del Banco en el programa que estimó probablemente surgirían en el futuro. En la revisión se resumieron enseñanzas de la evaluación y las suyas propias, incluyendo aquellas referentes al programa mismo (la necesidad de establecer un marco apropiado de seguimiento y evaluación), aquellas aplicables a otros GRPP (necesidad de coordinación y consulta eficaz con otros actores clave a nivel global y de país) y aquellas aplicables a otros instrumentos de ayuda (enseñanzas para proyectos que dependen de asociaciones público-privadas para la prestación de servicios).

Resumen

Hace poco tiempo, los evaluadores estuvieron intentando investigar los efectos de muchas intervenciones con un enfoque conceptual en un sector, un país o un tema. Con estos enfoques se intenta determinar cómo las intervenciones afectan una visión más general. Con los enfoques más complejos se examinan múltiples proyectos, programas o políticas para identificar interacciones y formas de compartir recursos. En esas evaluaciones se incluyen aquellas que son conjuntas, de programa de países, de programas sectoriales, temáticas y las de programas asociativos globales o regionales.

Referencias y lecturas adicionales

BAfD (Banco Africano de Desarrollo). 2005. *Morocco: Evaluation of Bank Assistance to the Education Sector.* http://www.oecd.org/dataoecd/36/19/37968449.pdf.

BAsD (Banco Asiático de Desarrollo). 2006a. *Indonesia: Capacity Building Project in the Water Resources Sector.* Project Performance Evaluation Report for Indonesia. http://www.adb.org/Documents/PPERs/INO/26190-INO-PPER.pdf.

———. 2006b. *Private Sector Development and Operations: Harnessing Synergies with the Public Sector.* http://www.oecd.org/dataoecd/6/59/39519572.pdf.

Bamberger, Michael. 2005. *IPDET Handbook for Evaluating Gender Impacts of Development Policies and Programs.* International Program for Development Evaluation Training, Ottawa.

Bamberger, Michael, Mark Blackden, Lucia Fort y Violeta Manoukian. 2001. "Integrating Gender into Poverty Reduction Strategies". En *The PRSP Sourcebook.* 335-74. Washington, DC: Banco Mundial. http://povlibrary.worldbank.org/files/4221_chap10.pdf.

Banco Mundial. 2001a. *Design Paper for a Multipartner Evaluation of the Comprehensive Development Framework.* Comprehensive Development Framework Secretariat, Washington, DC. http://www.worldbank.org/evaluation/cdf/cdf_evaluation_design_paper.pdf.

———. 2001b. *Engendering Development—through Gender Equality in Rights, Resources and Voice.* Washington, DC: Banco Mundial.

———. 2001c. *The PRSP Sourcebook.* http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPRS/0,,menuPK:384207~pagePK:149018~piPK:149093~theSitePK:384201,00.html.

———. 2005. *Evaluation of World Bank's Assistance to Primary Education.* Operations Evaluation Department, Washington, DC. http://www.worldbank.org/oed/education/evaluation_design.html.

———. 2007a. *Country Assistance Evaluation: CAE Retrospective.* Independent Evaluation Group, Washington, DC. http://www.worldbank.org/ieg/countries/cae/featured/cae_retrospective.html.

———. 2007b. *Global Program Review: Medicines for Malaria Venture.* Independent Evaluation Group, Washington, DC. http://lnweb18.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/d591aea3bbb897de852573130077a0cb?OpenDocument.

———. 2007c. *Impact Evaluations: Bangladesh Maternal and Child Health.* Independent Evaluation Group, Washington, DC. http://www.worldbank.org/oed/ie/bangladesh_ie.html.

———. 2007d. *Poverty Monitoring Systems.* http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPAME/0,,contentMDK:20203848~menuPK:435494~pagePK:148956~piPK:216618~theSitePK:384263,00.html.

———. 2007e. *PRSP Sourcebook.* Washington, DC. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPRS/0,,contentMDK:20175742~pagePK:210058~piPK:210062~theSitePK:384201,00.html.

———. 2007f. *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards.* Independent Evaluation Group, Washington, DC. http://siteresources.worldbank.org/EXTGLOREGPARPRO/Resources/sourcebook.pdf.

———. 2008. *CAE Methodology.* Independent Evaluation Group, Washington, DC. http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTCOUASSEVA/0,,contentMDK:21107046~menuPK:3098030~pagePK:64168445~piPK:64168309~theSitePK:2832047,00.html.

BMD (Banco Multilateral de Desarrollo), EC (Evaluation Cooperative Group) y WGEC (Working Group on Evaluation Criteria and Ratings for Public Sector Evaluation). 2003. *Good Practice Standards for Evaluation of MDB Supported Public Sector Operations.* http://www.ifc.org/ifcext/oeg.nsf/AttachmentsByTitle/MDB-ECG/US$FILE/MDB-ECG_Good-Practice.pdf.

Breier, Horst. 2005. *Joint Evaluations: Recent Experiences, Lessons Learned and Options for the Future.* Borrador del informe para the Development Assistance Committee Network on Development Evaluation, Organización para la Cooperación y el Desarrollo Económicos, París.

Chen, Jhaoying y Hans Slot. 2007. "Country-Led Joint Evaluation: Dutch ORET/MILIEV Programme in China". Presentación en la sexta reunión de la Development Assistance Committee Network on Development Evaluation, París, junio. http://www.oecd.org/dataoecd/63/28/38851957.ppt.

Compton, Donald W., M. Baizerman y S. H. Stockdill, eds. 2002. *The Art, Craft, and Science of Evaluation Capacity Building.* New Directions for Evaluation 93 (primavera) (publicación de la American Evaluation Association).

Danish Ministry of Foreign Affairs. 1999. *Evaluation Guidelines.* 2ª ed. Copenhague. http://www.um.dk/NR/rdonlyres/4C9ECE88-D0DA-4999-9893-371CB351C04F/0/Evaluation_Guidelines_1999_revised.pdf.

———. 2006. *Evaluation Guidelines.* Copenhague.

———. 2007. *Joint External Evaluation: The Health Sector in Tanzania 1999-2006.* http://www.oecd.org/dataoecd/53/46/39837890.pdf.

Duursma, M. 2004. *Child Labour in Scavenging, Tanzania, Dar es Salaam. Actualizada en agosto* 16, 2005. http://www.waste.nl/page/724.

"Evaluation of the Implementation of the Paris Declaration. Synthesis Report.: 2008, 26.

Freeman, Ted. 2007. "Joint Evaluations". Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Ottawa, junio-julio.

FNUAP (Fondo de Población de las Naciones Unidas). 2005. *State of World Population 2005.* http://www.unfpa.org/swp/2005/english/notes/index.htm.

Fullan, M. 1993. *Change Forces.* Londres: Falmer Press.

GTZ (Gesellschaft für Technische Zusammenarbeit). 2004. *National Monitoring of Sustainable Poverty Reduction Strategy Papers PRSPs.* Eschborn, Alemania. http://siteresources.worldbank.org/INTISPMA/Resources/Training-Eventsand-Materials/summary_MainReport.pdf.

Heath, John, Patrick Grasso y John Johnson. 2005. *World Bank Country, Sector, and Project Evaluation Approaches.* Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Ottawa, julio.

House, Ernest R. *Guiding Principles for Evaluators.* 1995. New Directions for Program Evaluation N° 66. San Francisco: Jossey-Bass.

Johnson, John. 2007. "Confronting the Challenges of Country Assistance Evaluation". Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Carleton University, Ottawa, junio 26-27.

Kusek, Jody Zall y Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System.* Banco Mundial, Washington, DC. http://www.oecd.org/dataoecd/23/27/35281194.pdf.

Mackay, Keith. 1999. *Evaluation Capacity Development: a Diagnostic Guide and Action Framework.* ECD Documento de trabajo 6 (enero), Banco Mundial, Washington, DC. http://lnweb18.worl-

dbank.org/oed/oeddoclib.nsf/a4dd58e444f7c61185256808006a0008/7f2c924e183380c5852567fc00556470?OpenDocument.

———. 2006. *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management*. Banco Mundial, Independent Evaluation Group, Washington, DC. http://siteresources.worldbank.org/INTISPMA/Resources/ecd_15.pdf.

———. 2007. *How to Build M&E Systems to Support Better Government.* Banco Mundial, Washington, DC. http://www.worldbank.org/ieg/ecd/docs/How_to_build_ME_gov.pdf.

Mertens, D. M. 1994. "Training Evaluators: Unique Skills and Knowledge". *New Directions for Program Evaluation* 62: 17-27.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). 1999. "Evaluating Country Programmes", Development Assistance Committee, Vienna Workshop, marzo 11-12. http://www.oecd.org/dataoecd/41/58/35340748.pdf.

———. 2002. *OECD Glossary of Key Terms in Evaluation and Results Based Management.* Development Assistance Committee, París.

———. 2005. Taller sobre "Challenging the Conventional Wisdom: The View from Developing Country Partners", Network on Development Evaluations, Development Assistance Committee, Nairobi, abril 20-21. http://www.oecd.org/dataoecd/20/44/34981186.pdf.

———. 2006. *DAC Evaluation Series: Guidance for Managing Joint Evaluations.* Development Assistance Committee, París. http://www.oecd.org/dataoecd/29/28/37512030.pdf.

OIM (Organización Internacional para las Migraciones). 2006. *IOM Evaluation Guidelines.* Office of the Inspector General. http://www.iom.int/jahia/webdav/site/myjahiasite/shared/shared/mainsite/about_iom/eva_techref/Evaluation_Guidelines_2006_1.pdf.

Porteous, Nancy L., Barbara J. Sheldrick y Paula J. Stewart. 1999. "Enhancing Managers' Evaluation Capacity: A Case Study for Ontario Public Heath". *Canadian Journal of Program Evaluation* (edición especial): 137-54. http://www.phac-aspc.gc.ca/php-psp/pdf/toolkit/enhancing_managers_evaluation%20_capacity%20_CJPE_1999.pdf.

WASTE. 2005. *Thematic Evaluation on Child Labour in Scavenging Africa, Asia, and Europe Assessment.* Gouda, Países Bajos. http://www.waste.nl/page/720.

Sitios web

Banco Mundial. *ADePT: Stata Software Platform for Automated Economic Analysis.* http://econ.worldbank.org/programs/poverty/adept.

———. *Core Welfare Indicators Questionnaire (CWIQ).* http://www4.worldbank.org/afr/stats/cwiq.cfm.

———. *Global and Regional Partnership Programs.* http://www.worldbank.org/ieg/grpp.

Organización Internacional del Trabajo. http://www.ilo.org.

Stufflebeam, D. L. 1999. *Evaluation Plans and Operations Checklist.* http://www.wmich.edu/evalctr/checklists/plans_operations.htm.

———. 2001. *Guiding Principles Checklist.* http://www.wmich.edu/evalctr/checklists/guiding_principles.pdf.

DIRIGIR

La acción debe culminar en la sabiduría.

—BHAGAVAD GITA

Capítulo 12. Administrar una evaluación

- Manejo de la matriz de diseño.
- Contratar la evaluación.
- Reglas y responsabilidades de los distintos actores.
- Administración de personal, tareas y presupuestos.

Capítulo 13. Presentación de resultados

- Elementos básicos de comunicación.
- Escritura de informes de evaluación para la audiencia.
- Uso de la información visual.
- Presentaciones orales.

001El camino.indd 413 08/11/2010 03:27:58 p.m.

Capítulo 14. Guiar al evaluador: ética, política, estándares y principios orientadores

- Comportamiento ético.
- Política y evaluación.
- Estándares de evaluación y principios orientadores.

001El camino.indd 414 08/11/2010 03:27:58 p.m.

CAPÍTULO 12

Administrar una evaluación

Las evaluaciones son empresas complicadas; mantener ocupada a la gente, cumplir con fechas límite y realizar un trabajo de alta calidad es una labor retadora. En el presente capítulo se discuten la forma como los evaluadores pueden planear, administrar, cumplir estándares de calidad y compartir resultados, de manera que quienes formulan las políticas utilicen sus evaluaciones para efectuar cambios.

Manejo de la matriz de diseño

La clave de las evaluaciones para el desarrollo exitosas es la planeación. Si la evaluación se planea con deficiencias, ninguna cantidad de análisis posterior –no importa cuán sofisticado pueda ser– la salvará. Según un adagio chino, un viaje de mil millas empieza con un paso. La probabilidad de llegar al destino aumenta mucho si el primer paso y los pasos siguientes llevan al viajero en la dirección correcta. Deambular sin un sentido claro de propósito o de dirección consume tiempo, energía y recursos, no solamente del viajero sino también del evaluador. También disminuye la posibilidad de llegar a algún sitio. Es, por consiguiente, sensato prepararse para un viaje, recopilando los mapas necesarios, estudiando rutas alternativas y haciendo estimados informados del tiempo, costos y riesgos que es probable se hayan de confrontar.

001El camino.indd 415 08/11/2010 03:27:58 p.m.

Como se mostró en los capítulos anteriores, la matriz de diseño de evaluación constituye una forma visual de mapear una evaluación. Se concentra en cada uno de los mayores componentes del diseño de una evaluación. La plantilla de la matriz no está labrada en piedra y puede adaptarse para las necesidades de la evaluación. Al igual que cualquier plan, probablemente pasará por un proceso iterativo de actualizaciones y revisiones. A medida que se adquiere nueva información, será necesario hacer correcciones a algunas ideas y enfoques.

Una buena matriz de diseño de evaluación no es lo único que se requiere para efectuar una buena evaluación. Los mejores diseños pueden dar como resultado una evaluación de baja calidad si, por ejemplo, personas sin las destrezas necesarias intentan implementarla, o la evaluación se entrega demasiado tarde para influir en una decisión crítica, o el presupuesto se agota durante la recopilación de los datos.

Contratar la evaluación

El director de la evaluación es responsable de identificar y remediar vacíos de destrezas en el personal disponible para realizar una evaluación. La oferta de personal con destrezas de evaluación puede ser escasa y la demanda alta. Por esta razón, las evaluaciones para el desarrollo se contratan con mucha frecuencia con empresas consultoras o consultores individuales. Algunas instituciones, por ejemplo, la Unión Europea, contratan la mayoría o el total de sus evaluaciones con especialistas. Para garantizar que su objetividad no se vea comprometida, los consultores no deberían haber participado anteriormente en el programa que se va a evaluar.

Utilizar contratistas tiene ventajas y desventajas. Las ventajas incluyen el hecho de que el contratista tiene conocimiento profundo del sector y el tipo de programa que se evalúa; habla el idioma local y tiene conocimiento local; y, si se utilizó un proceso competitivo, es el evaluador más efectivo en costos disponible. Las desventajas incluyen el hecho de que contratar es costoso (el proceso de ofertas puede costar más que el mismo contrato) y significa que no aumenta la capacidad interna.

Solicitud de propuestas: Invitación a cotizar para un proyecto.

Solicitud de propuestas

El proceso de contratar una empresa consultora, y algunas veces consultores individuales, se inicia desarrollando una *solicitud de propuestas* (SP). De acuerdo con la organización y las reglas y los procedimientos de los contratos, puede otorgarse el contrato con base en una sola fuente, mediante invitaciones a unos pocos contratistas para que envíen propuesta, o por competencia abierta. La propuesta ganadora es seleccionada por el administrador de la evaluación, o por un panel nombrado que haga su recomendación basado en criterios preestablecidos.

Hawkins (2005) sugiere que la SP incluya los elementos siguientes:

- Propósito de la evaluación.
- Antecedentes y contexto para el estudio.
- Requisitos básicos de información.
- Objetivos de la evaluación.
- Entregas requeridas.
- Marco de tiempo.
- Criterios para la selección de propuestas.
- Detalles del contrato para el administrador del proyecto.
- Fecha límite para las propuestas.
- Presupuesto y otros recursos.

La selección de la propuesta ganadora debe hacerla un panel de personas con conocimiento y experiencia en evaluaciones, conocimiento del área del programa, conocimiento de la cultura y propiedad de los hallazgos y sus usos. Este panel debe seleccionar la propuesta, utilizando los criterios de la SP. Deben mantenerse buenos registros sobre el proceso de selección.

Para seleccionar un consultor, Hawkins aconseja utilizar los criterios siguientes:

- ¿Se ha tratado adecuadamente la SP?
- ¿Existe una explicación detallada de cómo se implementará la evaluación?
- ¿Qué estrategia de comunicación y presentación de información se propone?
- ¿Hay evidencia de competencia? ¿Cuál es el historial de cumplimiento?
- ¿Cuál es el costo? ¿Se especifica en detalle?

Una vez contratado el consultor, el director de la evaluación tiene todavía responsabilidades con el contratista y la evaluación, en las cuales se incluyen las siguientes:

- Mantener claros las metas y objetivos.
- Mantener la posesión del estudio.
- Hacer seguimiento al trabajo y proporcionar retroalimentación oportuna.
- Tomar decisiones, a tiempo.
- Estar abierto a la negociación con el contratista si se requieren cambios al contrato.

Términos de referencia

Constituyen una declaración de antecedentes, objetivos y propósito de una evaluación; los deberes y responsabilidades individuales de los integrantes de un equipo de evaluación, y fechas límite para las entregas. Describen usualmente la evaluación general y fijan los acuerdos iniciales. Deben establecerse por acuerdo entre un cliente y un director de la evaluación, un director de la evaluación y un contratista, o un director de la evaluación y el personal que realizará ésta.

Términos de referencia (TR): Declaración de antecedentes, objetivos y propósito de una evaluación; los deberes y responsabilidades individuales de los integrantes de un equipo de evaluación; y fechas límite para las entregas.

Todos los integrantes del equipo de la evaluación deben tener un conocimiento claro y por escrito de sus responsabilidades específicas para la evaluación, que usualmente se redactan como TR.

En los TR debe definirse de manera realista lo que se necesita llevar a cabo en un período de tiempo definido.

Los TR pueden incluir la matriz de diseño, o el desarrollo de la matriz de diseño puede ser una de las primeras entregas. Pueden también reflejar acuerdos sobre ética y estándares en la realización y la presentación de información de la evaluación (*ver* capítulo 14, para una discusión sobre ética y estándares).

El proceso de desarrollar los TR generales es útil para asegurar que las partes interesadas queden incluidas en la discusión y en la toma de decisiones sobre los temas que se tratarán en la evaluación. Los TR establecen los lineamientos básicos, de modo que todos los participantes conozcan las expectativas para la evaluación (recuadro 12.1).

Según el *Glossary of Key Terms in Evaluation and Results-Based Management* (OCDE, 2002), los TR para una evaluación se presentan en un documento escrito que indica:

- El propósito de la evaluación.
- Los métodos que se utilizarán en ella.
- Los estándares contra los cuales se comparará el desempeño o se efectuarán los análisis.
- Los recursos y el tiempo asignados.
- Los requerimientos de presentación de información.

Los TR incluyen de manera usual:

- Un título corto y descriptivo.
- Una descripción del proyecto o programa.
- Las razones y expectativas de la evaluación.
- Una declaración del alcance y el enfoque de la evaluación (temas que serán tratados, preguntas que se espera responder).
- Identificación de la participación de las partes interesadas (quién participará, quién hará qué cosas, proceso de rendición de cuentas).
- Una descripción del proceso de evaluación (que se realizará)
- Una lista de entregas (plan de trabajo de la evaluación, informe preliminar, informe final, presentaciones).
- Identificación de las calificaciones necesarias (educación, experiencia, destrezas, capacidades requeridas).
- Proyección de costos basada en actividades, tiempo, número de personas, honorarios profesionales, viajes y otros costos.

Los TR se agregan a la matriz de diseño de la evaluación, no son un reemplazo de ella. Los TR se centran en responsabilidades. La matriz de diseño es un plan para realizar la evaluación.

001El camino.indd 418 08/11/2010 03:27:58 p.m.

Recuadro 12.1 *Consejos para escribir los términos de referencia para una evaluación*

Los consejos siguientes pueden ser útiles para escribir los términos de referencia para una evaluación:

- Expresar claramente los objetivos de la evaluación, e identificar los puntos siguientes:
  - Los temas generales y preguntas de evaluación preliminares que se tratarán.
  - Las partes interesadas clave y el uso esperado que le darán a la evaluación.
  - El enfoque general que se adoptará en la evaluación.
  - Los productos que se esperan de la evaluación, cuándo debe entregarse cada uno y cómo se utilizará.
  - La pericia requerida de los integrantes del equipo de evaluación.
  - Los arreglos de logística.
- No expresar simplemente los objetivos en términos técnicos o de procesos. Tener claridad en cuanto a cómo se espera que la evaluación ayude a la organización.
- Centrarse en los asuntos preliminares que deberán abordarse.
- Evitar seleccionar demasiadamente las preguntas. Es mejor examinar en una evaluación unos pocos asuntos en profundidad que examinar una amplia variedad de asuntos de manera superficial.

Fuente: adaptado de PNUD, 2006.

Papeles y responsabilidades de los distintos actores

En una evaluación trabajan muchas personas, que tienen diferentes capacidades y cumplen distintos papeles y responsabilidades. Los papeles principales son los del cliente (solicitante), las partes interesadas, el director de la evaluación y el evaluador (sea un contratista o el personal). El personal encargado de funciones limitadas incluye los analistas de datos, asesores, revisores pares y recopiladores de datos. Los papeles y responsabilidades de los participantes en la evaluación deben definirse y acordarse con claridad.

El cliente principal

Una evaluación puede tener muchos interesados pero de manera usual solamente un cliente principal que solicita y con frecuencia financia la evaluación. Los temas específicos que el cliente desea se examinen, el uso que el cliente planea darle a la evaluación y las correspondientes implicaciones de tiempo enmarcan generalmente la evaluación.

A los evaluadores se les alienta a reunirse pronto con el cliente para discutir los asuntos siguientes:

- La naturaleza y el contexto de la solicitud.
- La necesidad, el asunto o el interés que llevó a la solicitud para la evaluación.
- Las necesidades críticas de tiempo para los hallazgos de la evaluación.
- Las preguntas de más importantes que debe responder la evaluación.
- El programa y la frecuencia de las comunicaciones.

En la Key Evaluation Checklist, de Scriven (2007) se describe la importancia de reunirse con el cliente antes de iniciar la evaluación con el fin de identificar los detalles de la solicitud, como los ve el cliente, o estimular al cliente para que clarifique su concepción de los detalles. Scriven sugiere hacer preguntas como las siguientes:

- ¿Exactamente qué debería evaluarse?
- ¿Qué medida de contexto debe incluirse?
- ¿Debe el evaluador evaluar los efectos del programa como un todo o la contribución de cada uno de sus componentes? ¿Evaluará la teoría del cliente sobre cómo funcionan los componentes?
- ¿Debería el evaluador analizar el impacto en todos los aspectos pertinentes o solamente algunos aspectos?
- ¿Será la evaluación formativa, final y/o descriptiva?

Según se vio en el capítulo 4, el equipo de evaluación maneja su relación con el cliente principal, en parte desarrollando un plan de comunicaciones. Un plan de comunicaciones ayuda a los evaluadores a trabajar con el cliente, las partes interesadas clave, los medios y el público. En él se definen el porqué, el qué, el dónde y el cómo para la entrega y la recepción de información. También en él se fija un plan para comunicar, en el que usualmente se incluyen diversos medios para diferentes audiencias. El establecimiento de un plan de comunicaciones es especialmente importante si en la evaluación se incluye información sensible.

Los evaluadores consideran a menudo la comunicación como algo que sucede al final del proceso, cuando se ha producido el informe. En realidad, la comunicación y el desarrollo de una estrategia de comunicaciones deben tratarse por anticipado y refinarse durante el proceso. La comunicación debe utilizarse para dar soporte a la evaluación, mejorarla y mejorar su metodología, contribuir a obtener aceptación y promover el uso de los hallazgos.

Interesados

Según se definió en el capítulo 4, las partes interesadas son las personas y organizaciones distintas al cliente que tienen interés en la intervención. De manera usual, con aquellos afectados por una intervención, durante su duración o en años posteriores.

Las partes interesadas pueden ser internas o externas a la organización. Su papel en una evaluación puede ser desde identificar problemas y preguntas para la evaluación hasta suministrar datos.

En las evaluaciones participativas, las partes interesadas pueden asumir responsabilidades en varios niveles para la evaluación. En tales evaluaciones se pueden incluir las partes interesadas seleccionadas como parte del equipo de evaluación, incluyéndolos en el desarrollo de las preguntas, la recopilación de datos y su análisis. Es más probable que las partes interesadas apoyen la evaluación y actúen sobre los resultados y recomendaciones si participan en el proceso de evaluación. (Para una discusión de las evaluaciones participativas, *ver* el capítulo 5.)

Puede ser benéfico vincular a críticos del proyecto, programa o política motivo de la evaluación. Dependiendo del grado al que identifiquen temas importantes para la evaluación (que podrían incluso desacreditarla si no se tratan), la participación de críticos fortalece el proceso de la evaluación (CDC, 2001).

El *Participation Sourcebook* del Banco Mundial proporciona una orientación útil sobre la participación de las partes interesadas que es particularmente aplicable a las evaluaciones participativas (Banco Mundial, 1996). Una vez se ha identificado a las partes interesadas, debe inscribirse su participación. Los evaluadores han buscado trabajar con las partes interesadas mediante varios enfoques. Con mucha frecuencia se necesita medidas especiales para asegurar que tengan voz grupos que normalmente son excluidos del proceso de toma de decisiones. Para conseguir esto, los evaluadores deben primero organizar a los sin voz, arreglar su representación, sostener reuniones participativas con ellos y utilizar técnicas que permitan a los interesados ser escuchados en todos los niveles.

Para muchos interesados, alguien externo que trae ofertas de "evaluación participativa" puede parecer sospechoso. La experiencia con agencias públicas, servidores públicos, y proyectos de donantes puede haber creado impresiones negativas que deben rectificarse. La confianza puede lograrse compartiendo información, trabajando mediante intermediarios e iniciando pronto y ampliamente. Una forma de lograr confianza es la de compartir información sobre lo que se intenta con la evaluación, lo cual pueden hacer los evaluadores por medio de reuniones individuales de grandes grupos, por ejemplo, reuniones de la población. Durante estas reuniones los evaluadores pueden compartir información sobre las causas y los motivos de la evaluación. En la reunión los participantes tienen la oportunidad de expresar sus expectativas y preocupaciones. Si la evaluación es participativa, una vez lograda la confianza, puede invitarse a las partes interesadas a conformar sus propios comités para trabajar en la evaluación.

Una forma de abrir la comunicación con la comunidad académica y de investigación, ONG y ciudadanos interesados es publicar documentos de aproximación en sitios web y hacer la invitación a enviar comentarios. Desde luego, esto debe hacerse cuando todavía puedan tenerse en cuenta nuevos aportes y hacerse revisiones, y resulta más efectivo cuando los evaluadores publican respuestas que resumen los comentarios y acciones ejercidas (o razones para no incluir las sugerencias).

En algunas instancias la desconfianza es tan grande que puede requerirse que intermediarios llenen el vacío entre los evaluadores y las partes interesadas. En tales casos,

001El camino.indd 421 08/11/2010 03:27:58 p.m.

una de las partes que sea respetada por las partes interesadas puede utilizar su posición para reconciliar a los evaluadores y las partes interesadas.

Los evaluadores también deberían visitar a las partes interesadas, en especial a los afectados directamente por la intervención. Pueden presentarse y preguntar si las partes interesadas han oído sobre la evaluación y qué piensan de ella. Esta retroalimentación informal puede compararse con lo que el evaluador oye a través de otros medios. Sirve como forma de verificar la consistencia y verificar si hay sesgos.

Dar a las partes interesadas una voz en el diseño de la evaluación colectiva aumenta las posibilidades de resolver las diferencias y lograr un consenso sobre los distintos temas. Una relación temprana con las partes interesadas puede constituir así una medida para evitar conflictos. Para quienes perciben alguna pérdida para sí mismos en el diseño de la evaluación, la oposición directa puede ser la única instancia posible, y cuanto mayor la pérdida es más probable que sea más fuerte la oposición. Una vez se moviliza la oposición, es difícil –si no imposible– resolver el problema.

El director de la evaluación

Director de evaluación: Persona que dirige el diseño, preparación, implementación, análisis, presentación de información, divulgación y seguimiento de una evaluación.

Es la persona que dirige el diseño, la preparación, la implementación, el análisis, la presentación de información, la divulgación y el seguimiento de una evaluación. Puede dirigir varias evaluaciones al mismo tiempo y, en algunos casos, la misma persona puede dirigir y realizar la evaluación.

Sus responsabilidades pueden agruparse en tres etapas: preparación, implementación y seguimiento (UNFPA, 2007). La *preparación* incluye el trabajo anticipado de definir la evaluación y desarrollar las SP, los TR, la matriz de diseño de la evaluación y la preparación de la logística. Requiere que los directores hagan lo siguiente:

- Determinar el propósito e identificar los usuarios de los resultados de la evaluación.
- Determinar quién debe participar en el proceso de evaluación.
- Reunirse con las partes interesadas clave.
- Definir el alcance y el enfoque de la evaluación y la matriz de diseño, incluyendo las preguntas.
- Esbozar los TR para la evaluación e indicar el tiempo de duración de la evaluación.
- Identificar las destrezas y experiencias requeridas para realizar la evaluación.
- Supervisar el desarrollo y las pruebas preliminares de los instrumentos de recopilación de datos y la recolección de la información existente. Utilizar la selectividad y asegurar que las fuentes de información existentes sean confiables y de calidad suficientemente alta para producir resultados de evaluación significativos; la información recopilada debe ser manejable.

- Seleccionar, contratar e instruir preliminarmente a los evaluadores sobre el propósito de la evaluación, la matriz, el plan de trabajo y luego capacitarlos según las necesidades.
- Asegurar que la documentación y los materiales de antecedentes compilados se remitan con bastante anticipación a la evaluación, de manera que el equipo tenga tiempo de digerir los materiales.
- Supervisar el plan de visitas de campo.
- Asegurar la disponibilidad de fondos para la ejecución de la evaluación.

La *implementación* es la realización de la evaluación, incluyendo el análisis, el borrador y los informes. Esta requiere que los directores hagan lo siguiente:

- Asegurar que los evaluadores tengan acceso total a los archivos, informes, publicaciones y demás información pertinente.
- Seguir el proceso de la evaluación; proveer retroalimentación y orientación a los evaluadores durante la fase de implementación.
- Evaluar la calidad de los informes de evaluación y discutir fortalezas y limitaciones con los evaluadores, para asegurar que el borrador del informe cumpla con los TR, los hallazgos de la evaluación sean defendibles y las recomendaciones sean realistas.
- Arreglar una reunión con los evaluadores e interesados clave para discutir y comentar el borrador del informe.
- Aprobar el producto final y asegurar la presentación de los resultados de la evaluación a las partes interesadas.

La etapa final, el *seguimiento*, se refiere a evaluar el desempeño, divulgar los resultados, controlar el seguimiento de las recomendaciones, y efectuar un examen *post mortem*. Durante esta etapa, las responsabilidades del director incluyen las siguientes:

- Evaluar el desempeño de los evaluadores y registrarlo.
- Divulgar los resultados de la evaluación a las partes interesadas clave y otras audiencias.
- Promover la implementación de recomendaciones y el uso de los resultados de la evaluación en la programación actual y futura; hacer regularmente seguimiento para asegurar que se han seguido las recomendaciones.
- Dirigir el equipo en una revisión de aprendizaje para identificar qué se hizo bien y debería repetirse y qué, en retrospectiva, pudo haberse hecho en mejor forma.

Durante las tres etapas, el director de evaluación puede servir de facilitador en las reuniones del equipo, permitiendo que los participantes compartan opiniones e ideas. Como facilitador, el director es responsable de:

- Fijar una agenda.
- Ayudar al grupo a ceñirse a la agenda.

- Asegurar que se escuchen todas las opiniones.
- Supervisar un proceso para la toma de decisiones (consenso o votación).

El evaluador

Los evaluadores son las personas que hacen el trabajo principal en una evaluación. El número de evaluadores depende del tamaño y el alcance de la evaluación, el presupuesto y el número de personas disponibles.

El Programa de las Naciones Unidas para el Desarrollo (PNUD, 2006) identifica las siguientes características de un buen evaluador:

- Pericia en la materia específica.
- Conocimiento de temas de desarrollo básicos, especialmente los relativos a las metas principales, o la capacidad de tener la "visión general".
- Familiaridad con el negocio de la organización y la forma como se realiza.
- Destrezas de evaluación en diseño, recopilación de datos, análisis de datos y preparación de informes.
- Destrezas en el uso de tecnologías de información.

El Fondo de Población de las Naciones Unidas (UNFPA, 2007) identifica varias responsabilidades potenciales de los evaluadores:

- Hacer aportes sobre el diseño de la evaluación; hacer refinamientos y especificaciones a los objetivos y preguntas de la evaluación.
- Hacer la evaluación.
- Revisar la información y/o la documentación disponible.
- Diseñar o refinar instrumentos para recopilar información adicional según sea necesario; efectuar o coordinar la recopilación de información adicional.
- Efectuar visitas al sitio y entrevistas.
- En evaluaciones participativas, facilitar la participación de las partes interesadas.
- Suministrar informes o resúmenes regulares al director de la evaluación.
- Analizar y sintetizar la información, interpretar los hallazgos, desarrollar y discutir las conclusiones y recomendaciones y obtener enseñanzas.
- Participar en las discusiones del borrador del informe de la evaluación; corregir o rectificar errores factuales o malas interpretaciones.
- Guiar reflexiones y/o discusiones para facilitar la presentación de los hallazgos de la evaluación en un entorno de seminarios o talleres.
- Finalizar el informe de la evaluación y preparar una presentación de los resultados.

Muchas organizaciones intentan establecer criterios de competencia profesional para quienes se ocupan de la evaluación. También hay en camino discusiones para establecer la certificación para evaluadores. (Las competencias y la certificación para los evaluadores se examinan en el capítulo 15.)

001El camino.indd 424 08/11/2010 03:27:58 p.m.

Dirección de personas, tareas y presupuestos

Se han efectuado muchas comparaciones al describir la dirección de proyectos. Algunas se refieren a la dirección de proyectos y el malabarismo, porque el director debe mantener la vista en muchas cosas al mismo tiempo (personas, tareas, tiempo, presupuesto, calidad). Otras se hacen entre la dirección de un proyecto y la dirección de una orquesta, porque el director se mantiene a distancia y dirige a muchas personas competentes con diferentes destrezas.

Dirección de personas

Asegurar que un equipo de evaluación funcione bien en conjunto requiere varias destrezas. En esta sección se describen las destrezas de resolución de conflictos y formación de equipo que deben dominar los directores.

Los directores tienen también que manejar los temores de los "dueños" del programa. Algunas personas pueden estar preocupadas porque la evaluación se centre en lo negativo por ser eso lo que atrae la atención. Temen que algunos hallazgos negativos de la evaluación perjudiquen un programa que en general es bueno. También pueden surgir temores de que existan impactos del programa que no puedan medirse fácilmente, y esto produzca la conclusión errónea de que el programa es ineficaz, con consecuencias negativas para su financiación.

Una forma como los evaluadores pueden atender los temores de los dueños del programa es hacer que participen en la identificación de temas que podrían tratarse en la evaluación. Los evaluadores también deben ofrecerles oportunidad de revisar los planes de trabajo de la evaluación, los hallazgos y las recomendaciones.

Destrezas en resolución de conflictos

Los conflictos ocurren a menudo entre personas que trabajan en grupos. Las dos destrezas más necesarias para resolver los conflictos son las de comunicación y las de escucha. Las destrezas de comunicación importantes incluyen el uso de frases con "yo" en lugar de "usted". Las destrezas de escucha incluyen la de confirmar que el escucha ha entendido correctamente al interlocutor ("entiendo que usted dijo… ¿es correcto?").

No todos los conflictos deben terminar con un ganador y un perdedor. Los conflictos más constructivos terminan con "ganancia" para ambas partes (recuadro 12.2).

Destrezas de formación de equipos

Para trabajar en equipo se requieren algunas habilidades, entre ellas:

- *Escuchar*: Los integrantes del equipo deben desarrollar y utilizar buenas habilidades para escuchar, pues podrían incluso ser su activo más valioso.

Recuadro 12.2 *Consejos para resolver conflictos*

Los directores pueden resolver conflictos en varias formas:

- Reunir a quienes están en conflicto. Permitir que cada parte resuma brevemente su punto de vista, sin interrupción. Si una persona no permite que la otra termine o empieza a criticar, detenerla.
- Permitir a las personas expresar sus sentimientos, pues esto representa un avance importante en la resolución de conflictos.
- Solicitar a cada uno que describa las acciones que le gustaría que la otra persona ejerciera.
- Escuchar a ambas partes. Preguntarse a sí mismo si existe algo sobre la situación del trabajo que está causando el conflicto y, de ser así, buscar formas de modificar la situación para resolver el conflicto.
- No inclinarse hacia ninguna de las partes. Recordar a los participantes la meta de la evaluación, y esforzarse por encontrar un medio de ayudar a ambas partes a cumplir la meta.
- Esperar que los participantes trabajen para resolver sus disputas. Permitirles continuar reuniéndose para tratar el conflicto. Fijar un tiempo para revisar su progreso.

- *Preguntar:* los integrantes del equipo deben hacer preguntas para clarificar y elucubrar sobre lo que otros dicen.
- *Persuadir:* los integrantes del equipo deben intercambiar ideas, elucubrar, defender y repensar sus ideas.
- *Respetar:* los integrantes del equipo deben respetar las opiniones de los demás y alentar y apoyar sus ideas y esfuerzos.
- *Ayudar:* los integrantes del equipo deben ayudarse mutuamente, ofreciendo ayuda cuando sea necesario.

El manejo de un equipo de personas requiere algunas destrezas adicionales (recuadro 12.3).

Pueden utilizarse varias técnicas para obtener lo máximo posible de los equipos. La *sesión de intercambio de ideas* se utiliza para recopilar ideas de un grupo de personas en poco tiempo. Cada persona contribuye con una idea para una pregunta de la evaluación, la cual se escribe en un papelógrafo. Una persona da una idea y otra da otra. El facilitador recorre el grupo hasta que no surjan más ideas. La regla básica es que cada idea se escriba en el papelógrafo; ninguna idea se considera mala y no se discute ninguna de las ideas. En esta forma, se escuchan todas las ideas. El grupo en conjunto empieza entonces a identificar ideas comunes y se elabora una nueva lista.

Sesión de intercambio de ideas: Técnica utilizada para recopilar ideas de un grupo de personas en poco tiempo.

La *cartografía de grupos* es un proceso de facilitación de trabajo en grupos utilizado para solicitar ideas a todos los participantes, y agruparlas y clasificarlas según su importancia. Éste se empieza generando ideas, por medio de la sesión de intercambio

Recuadro 12.3 *Consejos para mejorar el equipo de trabajo*

Los consejos siguientes ayudan a los directores a mejorar la forma como trabajen juntos los integrantes del equipo:

- Comunicar los propósitos e importancia del trabajo especificado.
- Escuchar activamente, ayudar a otros a expresar su opinión mediante un parafraseo efectivo.
- Colaborar con los colegas y otras personas en prácticas indagatorias positivas, valorando la contribución particular de cada persona y su punto de vista.
- Apartar a un lado las predisposiciones personales en el afán de buscar respuestas a preguntas difíciles por el bien de las personas afectadas por el proyecto, el programa o la evaluación.
- Apartar a un lado las posturas defensivas frente a los hallazgos de la evaluación.

Fuente: NEIR TEC, 2004.

de ideas o la diagramación de afinidades. Cada idea se escribe en una tarjeta índice separada que se fija en la pared.

Cuando ya no surjan más ideas, se pide a los participantes examinar todas las ideas y arreglar las tarjetas en grupos que parezcan representar un concepto o un tema similar. Luego los participantes identifican el rótulo apropiado para cada concepto o tema. Entonces el facilitador dirige una discusión sobre la importancia de cada una y por qué es o no importante. En esta etapa pueden abandonarse algunos conceptos o temas relativamente no importantes, y también pueden abandonarse las ideas correspondientes a ellos o pasarse a otros temas. Cuando los participantes reciben los grupos señalados en la pared, pueden pedir también al facilitador trasladar otras ideas a otras categorías.

Cartografía de conceptos: Proceso de facilitación de trabajo en grupos utilizado para solicitar ideas a todos los participantes, y agruparlas y clasificarlas según su importancia.

Cuando los participantes estén satisfechos con los conceptos o temas y sus agrupaciones, el facilitador pasa a clasificarlos. A cada participante se le entregan varios rótulos adhesivos, dependiendo del número de conceptos o temas que el facilitador desee se seleccionen como importantes. Si son cinco, cada participante recibe cinco puntos de rótulos. Se pide a los participantes poner un punto al lado de cada uno de los cinco conceptos o temas que consideren más importantes. Luego el facilitador identifica los cinco conceptos o temas con mayor número de puntos y los clasifica según el número de puntos asignados. El proceso puede repetirse con las ideas correspondientes a cada concepto o tema, si se desea.

Administración de las tareas

La administración de las tareas parece algunas veces más fácil que la de personas, pero este proceso presenta sus propios problemas. Es muy importante mantenerse enfocado en la meta de la evaluación y las tareas principales.

Mapa de tareas: Tabla que identifica asignaciones, personas responsables y fechas de inicio y terminación.

Diagrama de Gantt: Gráfico utilizado para ilustrar la programación de un proyecto.

Un *mapa de tareas* puede ayudar con una lista de las asignaciones de las personas al lado de fechas de inicio y terminación (cuadro 12.2). También puede utilizarse un *diagrama de Gantt* (gráfico 12.1), que es una útil herramienta de seguimiento. La barra de cada tarea se llena para mostrar el progreso real contra lo planeado. El gráfico puede también mostrar la dependencia de una tarea con respecto a la terminación de otra, o un embotellamiento donde convergen muchas terminaciones de tareas. Muchos paquetes de *software* contienen plantillas de diagrama de Gantt.

Debe hacerse seguimiento a las actividades para asegurar que se completen las asignaciones según lo planeado. Si no se está realizando el progreso previsto, el evaluador debe identificar las barreras y determinar cómo eliminarlas. Es importante que los integrantes del equipo se sientan seguros para informar sobre problemas, pues a menudo es más fácil resolver un problema que se detecta con prontitud.

Aunque es importante tener un plan, lo es también ser flexible frente a obstáculos. Pueden efectuarse ajustes asignando más tiempo o recursos, o reduciendo el número de tareas incluidas.

Manejo de presupuestos

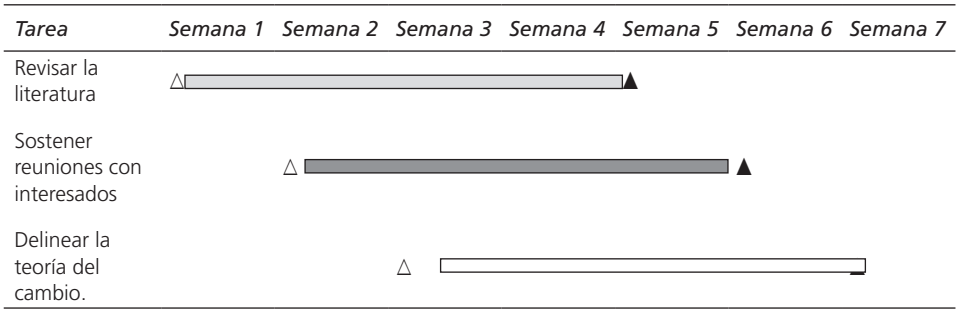
De acuerdo con Sanders (1983), el presupuesto de una evaluación incluye usualmente 10 categorías:

Cuadro 12.1 Muestra de mapa de tareas

Tarea	*Responsable(s)*	*Fecha de inicio*	*Fecha de terminación*
Revisar la literatura existente e identificar problemas relacionados internos y externos.	Ana, Miguel y Kabir	7/1	7/4
Revisar los documentos del programa o del proyecto (por ejemplo, documento de la junta, documentos de decisiones, informe de supervisión) específicos de la intervención.	Kabir	7/5	7/23
Programar y sostener reuniones con el cliente.	Ana	7/15	7/31
Identificar las partes interesadas clave y programar reuniones.	Kabir y Miguel	7/15	7/17
Escribir un resumen de la reunión con el cliente y de las decisiones tomadas.	Ana	8/1	8/3
Sostener reuniones con las partes interesadas y resumir los temas.	Ana y Miguel	8/5	8/15
Delinear el diseño inicial de la evaluación.	Ana	7/1	8/31

Fuente: los autores.

Gráfico 12.1 *Ejemplo de diagrama de Gantt*



Fuente: los autores.

- Salarios y beneficios del personal de evaluación.
- Honorarios de consultores.
- Gastos de viaje y gastos diarios (para personal y consultores).
- Comunicaciones (portes, llamadas telefónicas, etc.).
- Impresiones y copias.
- Procesamiento de datos.
- Materiales impresos.
- Equipos y suministros.
- Subcontratos.
- Gastos generales (instalaciones, utilidades).

Horn (2001) suministra una lista de verificación detallada para establecer el presupuesto de una evaluación. En esa lista se desglosan los elementos de costo correspondientes a las categorías presentadas. El director de la evaluación puede calcular el costo de cada categoría para estimar los costos totales de la evaluación.

Si el presupuesto para la evaluación resulta muy alto, puede ser necesario considerar algunas medidas de ahorro de costos desarrolladas por Fitzpatrick, Sanders y Worthen (2004):

- Utilizar voluntarios o trabajadores de bajo costo.
- Utilizar especialistas locales para recopilación de datos a fin de reducir los gastos de viaje.
- Capacitar personal menos costoso para tareas seleccionadas.
- Tomar en préstamo equipos, personas, materiales y suministros.
- Buscar contribuciones en especie en la organización en la que está empleado el evaluador externo.
- Reducir el alcance de la evaluación (quizá postergando algunas partes para el futuro).
- Utilizar medidas, datos o informes existentes.

- Utilizar recopilación económica de datos cuando pueda sacrificarse la precisión sin consecuencias importantes.
- Utilizar medios públicos para divulgar los resultados.
- Aumentar la eficiencia mediante una buena administración.

La proporción del presupuesto de un programa o un proyecto dedicada a la evaluación depende de muchos factores, incluyendo la visibilidad del proyecto o programa, el alcance planeado de la evaluación y el riesgo de reputación si la evaluación deja vacíos. La *Evaluation Toolkit* (1998) de la Kellog Foundation indica que un presupuesto de evaluación debería representar de un 5 a un 10% del presupuesto total de un proyecto. Otros especialistas en evaluación sugieren que se destine específicamente de un 1 a un 3% del costo del programa o del proyecto para la evaluación (el porcentaje es más pequeño para proyectos más grandes).

Administración de proyectos

La administración de un proyecto incluye el manejo de todos los aspectos del mismo, incluyendo el tiempo, el alcance, el costo y los recursos, el mismo tiempo. La administración de proyectos es un proceso de varias fases, y existen muchos modelos de administración de proyectos y varios programas de certificación en administración de proyectos. Michael Greer (2001), una autoridad en la administración de proyectos, desarrolló un modelo que hace énfasis en las acciones y los resultados. Su modelo divide la administración de un proyecto en cinco fases:

- Iniciación.
- Planeación.
- Ejecución.
- Control.
- Cierre.

Cada fase se divide en pasos, que Greer denomina *acciones*, e identifica 20 acciones distribuidas en las cinco fases (cuadro 12.2).

Resumen

La planeación es esencial para asegurar que las evaluaciones cumplan sus objetivos. Cada evaluador que trabaja en una evaluación debe conocer su papel y responsabilidades en lo que tiene relación con los demás que trabajan en la evaluación. Los términos de referencia ponen en claro estas responsabilidades por escrito, de modo que todos puedan revisarlas.

Al tratar los asuntos de personal, tiempo, tareas y costos, los directores deben tener presente la meta de la evaluación. Toda decisión debe basarse en cumplir la meta de la evaluación.

Cuadro 12.2 *Las 20 actividades de Greer para un administrador de proyecto*

Fase / acción	*Resultado del desempeño exitoso*
Fase de iniciación	
1. Demostrar la necesidad y la factibilidad del proyecto.	• Documento que confirma la necesidad de las entregas del proyecto y describe, en términos generales, los productos de entrega, los medios de crearlos, los costos de crearlos e implementarlos y los beneficios que se obtendrán al implementarlos.
2. Obtener la autorización para el proyecto.	• Toma de decisión de "seguir" o "no seguir" con el proyecto por parte del patrocinador. • Se asigna un director del proyecto, quien reconoce formalmente el proyecto y autoriza aplicar recursos a las actividades del proyecto. Se elabora un estatuto para el proyecto en el cual: a) se reconoce de manera formal el proyecto, b) un administrador externo lo remite al proyecto, y a nivel organizativo suficientemente alto para que pueda aplicar los recursos a las necesidades del proyecto, y c) autoriza al director del proyecto para aplicar los recursos a las actividades del proyecto.
3. Obtener autorización para la fase.	• El patrocinador toma la decisión de "seguir" o "no seguir", autorizando al director del proyecto para aplicar los recursos organizativos a las actividades de la fase particular. • Aprobación por escrito de la fase, que se crea y que: a) reconoce formalmente la existencia de la fase y b) la remite un administrador externo al proyecto, y a nivel organizativo suficientemente alto para que pueda satisfacer las necesidades del proyecto.
Fase de planeación	
4. Describir el alcance del proyecto.	• Declaración del alcance del proyecto. • Plan de gestión del alcance. • Estructura desglosada del trabajo.
5. Definir y poner en secuencia las actividades del proyecto.	• Lista de actividades que se realizarán. • Actualizaciones a la estructura desglosada del trabajo. • Diagrama de red del proyecto.
6. Estimar la duración de las actividades y los recursos requeridos.	• Estimado del tiempo requerido para cada actividad y supuestos relativos a cada estimado. • Declaración de requerimientos de recursos. • Actualizaciones a la lista de actividades.
7. Desarrollar un cronograma para el proyecto.	• Detalles de soporte, por ejemplo, uso de los recursos en el tiempo, proyecciones de flujo de caja, programas de orden/entrega, etc.
8. Estimar los costos.	• Estimados de costo para terminar cada actividad. • Detalles de soporte, incluyendo supuestos y restricciones. • Plan de manejo de costos, describiendo cómo se manejarán las variaciones en los costos.
9. Elaborar el plan de presupuesto y gastos.	• Costos básicos o presupuesto por fases de tiempo para medir y hacer seguimiento a los costos. • Plan de gastos, especificando cuánto se gastará en qué recursos y en cuánto tiempo.
10. Elaborar el plan formal de calidad (opcional).	• Plan de manejo de la calidad, incluyendo definiciones operativas. • Listas de verificación de la calidad.
11. Elaborar el plan formal de comunicaciones del proyecto (opcional).	• Plan de manejo de las comunicaciones, incluyendo estructura de recopilación, descripción de la información que se va a divulgar, cronogramas con listas de cuándo se producirá la información, método para actualizar el plan de comunicaciones.

(Continúa en la página siguiente)

(Continuación cuadro 12.2)

Fase / acción	*Resultado del desempeño exitoso*
12. Organizar y conseguir el personal.	• Asignación de funciones y responsabilidades. • Plan de personal. • Diagrama organizativo, con detalles apropiados. • Personal del proyecto. • Directorio del equipo del proyecto.
13. Identificar los riesgos y el plan de respuestas (opcional).	• Documento con descripción de riesgos potenciales, incluyendo fuentes, síntomas y formas de tratarlos.
14. Planear y adquirir los recursos externos (opcional).	• Plan de manejo de adquisiciones, describiendo la contratación de contratistas. • Declaración de trabajos o de requerimientos, describiendo el elemento (producto o servicio) que se va a adquirir. • Documentos de propuestas (ofertas). • Criterios de evaluación (medio de calificación de las propuestas de contratistas). • Contrato con uno o más proveedores de bienes o servicios.
15. Organizar el plan del proyecto.	• Plan integral del proyecto que agrupa los productos de las actividades anteriores de la planeación del proyecto.
16. Cerrar la fase de planeación del proyecto.	• Plan del proyecto que ha sido aprobado, por escrito, por parte del patrocinador ("luz verde" para iniciar los trabajos del proyecto).
17. Revisar el plan del proyecto y replantearlo si se requiere.	• Confianza en que los planes detallados de ejecución de una fase particular son todavía precisos, y efectivamente se lograrán los resultados planeados.
Fase de ejecución	
18. Ejecutar las actividades del proyecto.	• Resultados de los trabajos (productos de entrega). • Solicitudes de cambios (con base en la expansión o la contracción del proyecto). • Informes periódicos de progreso. • Evaluación y mejoramiento (si es necesario) del desempeño del equipo. • Solicitud de ofertas/propuestas para productos de entrega, selección de contratistas (proveedores) y restablecimiento de contratos. • Administración de contratos para lograr los resultados del trabajo deseados.
Fase de control	
19. Controlar las actividades del proyecto.	• Decisión de aceptar los productos de entrega inspeccionados. • Acciones correctivas, por ejemplo, reelaboración de productos de entrega, ajustes a procesos, etc. • Actualizaciones del plan y alcance del proyecto. • Lista de enseñanzas adquiridas. • Mejor calidad. • Listas completas de verificación de la evaluación (si se aplica).
Fase de cierre	
20. Cerrar las actividades del proyecto.	• Aceptación formal, documentada por escrito, de que el patrocinador aceptó el producto de esta fase o actividad. • Aceptación formal de los productos de los contratistas y actualización de los archivos de contratistas. • Actualización de los registros del proyecto preparados para su archivo. • Plan de seguimiento y/o no intervencionista de los productos de trabajo.

Fuente: Greer, 2001. Reimpreso con autorización.

001El camino.indd 432 08/11/2010 03:27:59 p.m.

Actividades del capítulo 12

Ejercicio de aplicación 12.1: términos de referencia

Lea los siguientes términos de referencia y siga las instrucciones que aparecen a continuación de ellos.

Por sugerencia de la Organización Mundial de Comercio (OMC), el grupo de varias agencias que maneja el programa del Marco Integrado solicitó al Banco Mundial asumir la dirección para realizar una revisión de la implementación del programa en los dos últimos años. El Marco Integrado, una empresa conjunta de varias agencias, busca ayudar a los países menos desarrollados a aprovechar las oportunidades que ofrece el sistema del comercio internacional. Ofrece asistencia relativa al comercio y seminarios sobre las reglas de la OMC para el mejoramiento de los puertos y atracaderos. La forma como funciona es ayudando a países individuales a identificar sus necesidades, y luego desarrollando un programa de asistencia solicitada que remite a una reunión de mesa redonda para el apoyo de donantes.

La revisión debe cubrir los temas siguientes:

- Percepciones de los objetivos del Marco Integrado (obtenidos mediante exploración de las opiniones de las partes participantes).
- Revisión de la asistencia relacionada con el comercio (formación de instituciones, formación de capacidad humana y empresarial e infraestructura).
- Consideraciones sobre políticas, incluyendo la ampliación del Marco Integrado y el entorno de políticas comerciales y macroeconómicas.
- Administración del Marco Integrado.
- Recomendaciones para el futuro.

Al cubrir estos temas, el consultor debe evaluar la pertinencia de las operaciones del Marco Integrado con respecto a sus objetivos. Debe evaluar también la efectividad en costos del Marco en el logro de sus objetivos, y la eficacia de la coordinación entre las agencias centrales que supervisan el Marco, las mesas redondas y otras actividades.

Se espera que el consultor examine la documentación sobre la implementación, realice entrevistas con el personal operativo de todas las agencias participantes, y busque a fondo las revisiones de los representantes de los países menos desarrollados y de los representantes del gobierno, y las empresas en por lo menos dos países menos desarrollados que se hayan beneficiado con el Marco Integrado, uno de los cuales debe ser de África (se proponen Botsuana y Uganda). Debe consultar, además, a representantes de los principales donantes. El informe debe tener unas 20 páginas sin incluir los anexos que se requieran.

Trabajando en parejas, de ser posible, revisar y criticar los términos de referencia y responder las preguntas siguientes:

1. ¿Los términos de referencia incluyen todos los elementos necesarios?
2. ¿Qué elementos están completos?
3. ¿Qué elementos pueden mejorarse?

Ejercicio de aplicación 12.2: ¿Está usted listo para ser director?

Lea la lista de características de un director (Reh, 2007). Identifique las capacidades que tiene y las que necesita mejorar.

Como persona

- Tiene confianza en sí mismo y en sus capacidades. Está contento con quien es, pero está aún aprendiendo y mejorando.
- Es algo extrovertido. No tiene que ser el alma de la fiesta, pero no puede quedarse sin pareja. La administración es una habilidad sobre personas; no es trabajo para alguien que no goza con la gente.
- Es honesto y franco. Su éxito depende en gran parte de la confianza de otros.
- Usted es un "incluyente", no un "excluyente". Atrae a otros hacia lo que hace y no excluye a otros porque carezcan de ciertos atributos.
- Usted tiene presencia. Los administradores deben dirigir. Los dirigentes eficaces tienen una cualidad propia que hace que la gente los note cuando entran a un recinto.

En el puesto

- Usted es consistente, pero no rígido; cumplidor, pero puede cambiar de opinión. Toma decisiones pero acepta con facilidad aportes de otros.
- Es un poco alocado, de pensamiento innovador, intenta cosas nuevas y, si fracasan, admite el error pero no se disculpa por haber intentado.
- No tiene temor de "hacer los cálculos". Hace planes y cronogramas y trabaja en consecuencia con ellos.
- Usted es ágil y puede cambiar rápidamente los planes, pero no es caprichoso.
- Considera la información como herramienta que debe utilizarse, no como un poder para acaparar.

Referencias y lecturas adicionales

Banco Mundial. 2006. *The World Bank Participation Sourcebook.* Washington, DC: Banco Mundial. http://www.worldbank.org/wbi/sourcebook/sbhome.htm.

Bemelmans Videc, M. L., R. C. Rist y E. Vedung. 1997. *Sticks, Carrots y Sermons: Policy Instruments and Their Evaluation.* Piscataway, NJ: Transaction Publishers.

Billson, Janet Mancini. 2004. *The Power of Focus Groups: A Training Manual for Social, Policy y Market Research: Focus on International Development.* Burlington, RI: Skywood Press.

CDC (U.S. Centers for Disease Control and Prevention). 2001. *Introduction to Program Evaluation.* http://www.cdc.gov/tobacco/tobacco_control_programs/surveillance_evaluation/evaluation_manual/00_pdfs/Chapter1.pdf.

Chelimsky, E. 1987. "The Politics of Program Evaluation". *Social Science and Modern Society* 25: 24-32.

De Regt, Jacomina. 1996. "Mozambique: Country Implementation Review". En *World Bank Participation Sourcebook,* 83-88. Washington, DC: Banco Mundial. http://www.worldbank.org/wbi/sourcebook/sb0211.pdf.

ESRC (Economic and Social Research Council). 2007. "Welcome to ESRC Today. Top Ten Tips". http://www.esrc.ac.uk/ESRCInfoCentre/Support/Communications_toolkit/communications_strategy/index.aspx.

Feuerstein, M. T. 1986. *Partners in Evaluation: Evaluating Development and Community Programs with Participants.* Londres: MacMillan, en asociación con Teaching Aids at Low Cost.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* Nueva York: Pearson Education.

FNUAP (Fondo de Población de las Naciones Unidas). 2007. "Tool N° 5: Planning and Managing an Evaluation". *Programme Manager's Planning, Monitoring and Evaluation Toolkit.* Nueva York. http://www.unfpa.org/monitoring/toolkit/5managing.pdf.

Greer, Michael. 2001. "20 Key Project Manager Actions and Results". En *The Project Manager's Partner.* Amherst, MA: HRD Press. http://www.michaelgreer.com/20-actns.htm.

———. 2008. *Michael Greer's Project Management.* http://www.michaelgreer.com.

Hawkins, Penny. 2005. "Contracting Evaluation". Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Ottawa, junio 30-julio 1.

Horn, Jerry 2001. *A Checklist for Developing and Evaluating Evaluation Budgets.* http://www.wmich.edu/evalctr/checklists/evaluationbudgets.pdf.

Kellogg Foundation. 1998. *Evaluation Handbook.* Battle Creek, MI. http://www.wkkf.org/Pubs/tools/Evaluation/Pub770.pdf.

King, Jean A., Laurie Stevahn, Gail Ghere y Jane Minnema. 2001. "Toward a Taxonomy of Essential Evaluator Competencies". *American Journal of Evaluation* 222: 229-47. http://www.nbowmanconsulting.com/Establishing percent20Essentialpercent20Programpercent20Evaluator percent20Competencies.pdf.

Kirkhart, K. E. 2000. "Reconceptualizing Evaluation Use: An Integrated Theory of Influence". En *The Expanding Scope of Evaluation Use,* ed. V.J. Caracelli y H. Preskill, 5-24. New Directions for Evaluation N° 88. San Francisco: Jossey-Bass.

Kusek, Jody Zall y Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System.* Washington, DC: Banco Mundial.

Lawrence, J. 1989. "Engaging Recipients in Development Evaluation: The 'Stakeholder' Approach". *Evaluation Review* 13 (3): 243-56.

Leeuw, Frans. 1991. "Policy Theories, Knowledge Utilization and Evaluation".

OECD World Forum on Statistics: Knowledge and Policy, Vol. 4, 73-91. Organización para la Cooperación y el Desarrollo Económicos, París.

McNamara, C. 1999. *Checklist for Program Evaluation Planning.* http://www.managementhelp.org/evaluatn/chklist.htm.

Muir, Edward. 1999. "They Blinded Me with Political Science: On the Use of Non-Peer Reviewed Research in Education Policy". *Political Science and Politics,* 32 (4): 762-64.

NCSTE (Chinese National Centre for Science and Technology Evaluation) e IOB (Policy and Operations Evaluation Department , los Países Bajos). 2006. *Country-Led Joint Evaluation of the ORET/MILIEV Programme in China.* Amsterdam: Aksant Academic Publishers.

NEIR TEC (Northeast and the Islands Regional Technology in Education Consortium). 2004. *Gathering Together and Planning: Exploring Useful Skills for Educators to Develop Through Collaborative Evaluation*. http://www.neirtec.org/evaluation/PDFs/Gathertogether3.pdf.

OCDE (Organización para la Cooperación y el Desarrollo Económicos). s.f. *DAC Evaluation Quality Standards for Test Phase Application.* Development Assistance Committee, París. http://www.oecd.org/dataoecd/30/62/36596604.pdf.

———. 2002. *OECD Glossary of Key Terms in Evaluation and Results-Based Management.* Development Assistance Committee, París.

Patton, Michael Q. 1977. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research". En *Using Social Research in Public Policy Making,* ed. C. H. Weiss, 141-64. Lexington, MA: Lexington Books.

———. 1997. *Utilization-Focused Evaluation.* 3ª ed. Thousand Oaks, CA: Sage Publications.

———. 2005. "Qualitative Methods and Analysis for Development Evaluation". Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Ottawa, junio 27-29.

Reh, F. John. s.f. "How to Be a Better Manager". http://management.about.com/cs/midcareermanager/a/htbebettermgr.htm.

———. 2007. "Management Tips". http://management.about.com/cs/generalmanagement/a/mgt_tips03.htm.

Rist, Ray C. y N. Stame. 2006. *From Studies to Streams: Managing Evaluativo Systems.* Piscataway, NJ: Transaction Publishers.

Rossi, Peter Henry, Howard E. Freeman y Mark W. Lipsey. 1999. *Evaluation: A Systematic Approach.* Thousand Oaks, CA: Sage Publications.

Rutman, L. 1980. *Planning Useful Evaluations: Evaluability Assessment.* Thousand Oaks, CA: Sage Publications.

Sanders, James R. 1983. "Cost Implications of the Standards". En *The Cost of Evaluation,* eds. M. C. Alkin y L. C. Solman, 101-17. Thousand Oaks, CA: Sage Publication.

Schwartz, R. 1998. "The Politics of Evaluation Reconsidered: A Comparative Study of Israeli Programs". *Evaluation* 4: 294-309.

Schwartz, R. y J. Mayne. 2004. *Quality Matters: Seeking Confidence in Evaluating, Auditing y Performance Reporting.* Piscataway, NJ: Transaction Publishers.

Scriven, Michael. 2007. "Key Evaluation Checklist". http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf/.

Stufflebeam, Daniel L. 1999. "Evaluations Plans and Operations Checklist". http://www.wmich.edu/evalctr/checklists/plans_operations.pdf.

Tilley, Nick. 2004. *Applying Theory-Driven Evaluation to the British Crime Reduction Programme: The Theories of the Programme and of Its Evaluations.* Thousand Oaks, CA: Sage Publications.

Weiss, Carol H. 1973. "Where Politics and Evaluation Research Meet". *Evaluation* 1: 37-45.

———. 2004. *Identifying the Intended Uses of an Evaluation.* International Program for Development Evaluation Training (IPDET); presentación, Ottawa, julio. http://www.idrc.ca/ev_en.php?ID=58213_201&ID2=DO_TOPIC.

Wholey, J. S. 1979. *Evaluation: Promise and Performance.* Washington, DC: Urban Institute.

———. 1994. "Assessing the Feasibility and Likely Usefulness of Evaluation". En *Handbook of Practical Program Evaluation,* eds. Joseph S. Wholey, Harry P. Hatry y Kathryn E. Newcomer, 15-39. San Francisco: Jossey-Bass.
Widavsky, A. 1972. "The Self-Evaluating Organization". *Public Administration Review* 32: 509-20.

Sitios web

Canadian Evaluation Society. *Professional Designation Project.* http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=EN.
CIID (Centro Internacional de Investigaciones para el Desarrollo). 2004. *Evaluation Planning in Program Initiatives.* http://web.idrc.ca/uploads/user-S/108549984812guidelineweb.pdf.
Conflict Resolution Information Source. http://www.crinfo.org/index.jsp.
Conflict Resolution Network. http://www.crnhq.org/.
EIB (European Investment Bank Group). 2008. "What Is the EIB Group?" http://www.eib.org/attachments/general/whatis_eibgroup_2008_en.pdf.
ESRC (Economic and Social Research Council). "ESRC Society Today: Communication Strategy". http://www.esrc.ac.uk/ESRCInfoCentre/Support/Communications_toolkit/communications_strategy/index.aspx.
Europe Aid Co-operation Office. 2005. *Evaluation Methods.* http://ec.europa.eu/europeaid/evaluation/methodology/index_en.htm.
———. 2006. *Methodological Bases.* http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_en.htm.
European Commission. 2008. *Evaluation Network.* http://ec.europa.eu/regional_policy/sources/docgener/evaluation/tech_en.htm.
Evalnet. 2000. http://www.evalnet.co.za/services/.
FNUAP (Fondo de Población de las Naciones Unidas). 2004. *Programme Manager's Planning, Monitoring and Evaluation Toolkit.* http://www.unfpa.org/monitoring/toolkit/5managing.pdf.
IBSTPI (International Board of Standards for Training, Performance and Instruction). *Competencies for Internal Staff and External Consultants Conducting Evaluations in Organizational Settings.* http://www.ibstpi.org/Competencies/evaluatorcompetencies.htm.
MSH (Management Sciences for Health) y UNICEF (Fondo de las Naciones Unidas para la Infancia). "Quality Guide: Stakeholder Analysis". En *Guide to Managing for Quality.* http://bsstudents.uce.ac.uk/sdrive/Martin percent20Beaver/Week percent202/Quality percent20Guide percent20-percent20Stakeholder percent20Analysis.htm.
PNUD (Programa de las Naciones Unidas para el Desarrollo). 2006. *Planning and Managing an Evaluation Website.* http://www.undp.org/eo/evaluation_tips/evaluation_tips.html.
Treasury of Board of Canada Secretariat. 2005. *Improving the Professionalism of Evaluation.* http://www.tbssct.gc.ca/eval/dev/Professionalism/profession_e.asp.
Weiss, Carol. *Identifying the Intended Use(s) of an Evaluation.* International Development Research Centre. http://www.idrc.ca/ev_en.php?ID=58213_201&ID2=DO_TOPIC.
Western Michigan University Evaluation Center. *Evaluation Checklists.* http://www.wmich.edu/evalctr/checklists/checklistmenu.htm.
———. *The Checklist Project.* http://evaluation.wmich.edu/checklists.

Ejemplos de términos de referencia

BAsD (Banco Asiático de Desarrollo). 2008. *Model Terms of Reference Diagnostic City Water Assessment* (Links). http://www.adb.org/Water/tools/City-Water-Assessments.asp.

CIDA (Canadian International Development Agency). *Model for Evaluation Terms of Reference.* http://www.acdicida.gc.ca/INET/IMAGES.NSF/vLUImages/Performancereview4/US$file/tor_sample_text.pdf.

FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). *FAO odel Terms of Reference for a Joint Evaluation Mission.* http://www.fao.org/pbe/pbee/common/ecg/233/en/torseng.doc.

PNUD (Programa de las Naciones Unidas para el Desarrollo). 2002. *Handbook on Monitoring and Evaluating for Results.* http://www.undp.org/eo/documents/HandBook/ME-HandBook.pdf.

CAPÍTULO 13

Presentación de resultados

Una vez que se hayan completado en gran parte la recopilación y el análisis de los datos, es tiempo de compartir los resultados preliminares y hacer planes para comunicar los resultados finales. Compartir lo aprendido es una de las partes más importantes de una evaluación. Es una condición previa crítica para efectuar el cambio. La presentación de los resultados puede hacerse por escrito (por medio de memorandos y de informes) o de manera oral (mediante resúmenes y presentaciones).

Ensamblar una estrategia de comunicación

Una evaluación que no se utilice como información para la toma de decisiones tiene poco valor. Cuando se diseña una evaluación, es crucial tener presente la meta de suministrar información útil para las partes interesadas que conduzca a tomar decisiones. En la evaluación no se busca obtener conocimiento por el conocimiento en sí, y ésta es una diferencia básica entre la evaluación y la investigación. Es esencial que los resultados de una evaluación se comuniquen con claridad y precisión, de manera que permita que la audiencia haga el mejor uso de ella.

001El camino.indd 439 08/11/2010 03:27:59 p.m.

Una estrategia de comunicación es un componente esencial de la evaluación para el desarrollo. A través de todo este texto hemos hecho énfasis en que la buena comunicación empieza en el puro principio y continúa durante toda la evaluación. No es una actividad que tenga lugar únicamente al final de una evaluación. El cliente y los principales interesados deben participar no solamente en la planeación de la evaluación, sino también en el desarrollo del proceco y los medios para comunicar sus hallazgos y solicitar retroalimentación.

La estrategia de comunicación debe desarrollarse desde el mismo principio. En ella debe identificarse:

- Quién necesita recibir información sobre la evaluación.
- Qué información se necesita.
- En qué formato debe estar la información.
- Cuándo debe suministrarse la información.
- Quién es responsable de suministrar la información.

Una buena estrategia de comunicación cubre todas las fases y productos de la evaluación. Utiliza varias formas de comunicación adecuadas a las necesidades de interesados particulares.

La estrategia de comunicación para una evaluación podría empezar con una discusión exploratoria de los temas con quien solicitó la evaluación, seguida de discusiones con grupos de interesados locales (cuadro 13.1). La discusión con el solicitante puede estar seguida de un resumen formal sobre el diseño preliminar para la evaluación. Los grupos de interesados locales pueden recibir una notificación por correo electrónico de que el diseño preliminar se encuentra en un sitio web para recibir comentarios. Las comunicaciones informales son valiosas durante la evaluación. Por ejemplo, los clientes y las partes interesadas pueden mantenerse informados del progreso de la evaluación mediante llamadas telefónicas o correo electrónico.

Los resultados finales pueden divulgarse a las diferentes partes interesadas de distintas formas. Por ejemplo, un donante podría recibir una sesión informativa profunda sobre los hallazgos de la evaluación, y luego un informe final formal. El personal local del programa podría recibir un informe general y una sesión informativa.

Para el informe final, los evaluadores pueden utilizar sesiones informativas, informes y presentaciones escritos. Los comunicados de prensa pueden utilizarse para divulgar información a una audiencia más amplia. Si se planea un comunicado de prensa o una conferencia de prensa, debe discutirse su oportunidad y manejo con los principales interesados.

Es importante incluir un proceso de retroalimentación para reunir a interesados y evaluadores para que discutan los hallazgos, percepciones, acciones alternativas y pasos que se van a seguir. Si se planea sostener grandes discusiones de grupo, los evaluadores deberían identificar los problemas para comunicar simultáneamente los resultados a grupos diferentes de interesados.

Aun los escritores de informes técnicos de evaluaciones conceden que no pueden ser la forma más eficaz de comunicar los hallazgos de la evaluación (Torres, Preskill

Cuadro 13.1 *Ejemplo de lista de verificación sobre la estrategia de comunicación para las fases preliminar y de diseño de la evaluación*

Audiencia	*Acción*	*Forma de comunicación*	*Responsable*	*Fecha límite*
Trabajo preliminar				
Cliente	Discutir los temas del programa y las necesidades de tiempo.	Reunión	Líder del equipo	6/1
Organización no gubernamental nacional y local	Discutir los temas del programa.	Reuniones	Integrante del equipo B	6/5
Personal del programa	Discutir los temas del programa.	Reuniones	Integrante del equipo C	6/11
Funcionarios del gobierno local	Discutir los temas del programa.	Reunión	Integrante del equipo B	6/10
Junta asesora	Identificar y enviar cartas de invitación.	Correo electrónico	Integrante del equipo A	6/14
	Planear y sostener reunión preliminar sobre los temas.	Reunión de junta asesora	Integrante del equipo B	6/25
Comunidad de evaluación para el desarrollo	Invitar a comentarios sobre los temas	Noticias por correo electrónico sobre sitio web abierto de la evaluación	Integrante del equipo C	6/25
Diseño				
Junta asesora	Revisar y discutir el diseño preliminar.	Reunión de junta asesora	Integrante del equipo A	7/15
	Proporcionar el diseño final.	Correo electrónico	Integrante del equipo A	7/30
Cliente	Parte final del diseño	Sesión informativa breve oral	Líder del equipo	7/22
Comunidad de evaluación para el desarrollo	Revisar el diseño preliminar.	Sitio web (seguimiento de comentarios)	Integrante del equipo B	7/31

Fuente: los autores.

y Piontek, 1997). Los evaluadores necesitan utilizar herramientas alternativas de comunicación, por ejemplo folletos, videos, documentos de exposición de problemas y resúmenes cortos con ilustraciones y gráficos para divulgar sus mensajes (recuadro 13.1).

Redactar informes de evaluación

El propósito de un informe es comunicarse con los lectores, para lo cual vale la pena tener en cuenta los consejos siguientes:

Recuadro 13.1 *Uso de técnicas innovadoras de comunicación para aumentar el interés en los resultados de la evaluación*

Para divulgar sus resultados a las partes interesadas, Lawrenz, Gullickson y Toal (2007) utilizaron varias técnicas innovadoras.

1. Escribieron una narrativa de caso discursiva, que divulgaron por medio de varios canales de salida, incluyendo Internet, para satisfacer las necesidades de su audiencia. La narrativa describía visitas de equipos de evaluadores a los sitios de 13 proyectos, y se tenían en cuenta no solamente las necesidades del cliente, sino también las de otras personas que podrían beneficiarse con los hallazgos.
2. Organizaron los resultados de la evaluación en nueve documentos de exposición de problemas (sobre colaboración, divulgación, desarrollo de materiales, desarrollo profesional, mejoramiento de programa, contratación y retención, sostenibilidad, comités asesores y evaluación) y los publicaron en su sitio web. Estos documentos sintetizaron los informes de visita a los sitios, los datos de las encuestas y la investigación previa por problema.
3. Para atraer el interés en los documentos de exposición de problemas crearon un folleto a manera de rompecabezas, que atrajo tanto interés hacia los documentos que publicaron el juego completo de documentos en una revista, divulgando los resultados a una audiencia más amplia que la planeada originalmente.
4. Colocaron en línea su manual de visitas a los sitios, con los procedimientos utilizados para efectuar las visitas. Después de publicarlo en su página web encontraron que muchas organizaciones y otros investigadores se interesaron en él.
5. Desarrollaron un folleto con pasos de acción básicos para la sostenibilidad del programa que estudiaron, el cual colocaron en el sitio web en formato pdf. Este folleto fue bien recibido recientemente por la comunidad de evaluación, porque era fácil de usar y suministraba información pertinente sobre el mejoramiento de los proyectos.
6. Sostuvieron una videoconferencia para discutir en profundidad algunas de sus ideas.
7. Colocaron hipervínculos entre la videoconferencia y el sitio web. La información del sitio incluyó vínculos a información sobre los autores del estudio, presentadores clave de la videoconferencia y documentos de soporte y materiales de video. Toda esta información fue publicada en el sitio web y también en un disco compacto.

- Tenga presente tanto el propósito como la audiencia al escribir el informe. Averigüe todo lo que le sea posible sobre la audiencia y escriba el informe de tal manera que se ajuste para llegar a ella.
- Utilice palabras:
  - Sencillas.
  - Activas.
  - Positivas.
  - Familiares.
  - Sensibles culturalmente.

- Evite al máximo abreviaturas y siglas.
- Limite la información de antecedentes a la necesaria para introducir el informe y hacer claro el contexto de éste. El contexto adicional puede incluirse como anexo si es necesario.
- Suministre suficiente información sobre el diseño y los métodos de la evaluación, de modo que los lectores tengan confianza en la credibilidad del informe, pero reconozcan sus limitaciones. Advierta a los lectores que algunas formas de interpretar los hallazgos pueden no ser válidas. Nuevamente, los detalles pueden colocarse en un anexo.
- Escriba un resumen ejecutivo (se describe a continuación).
- Organice el material del cuerpo del informe en secciones que traten los temas principales, o respondan las preguntas principales de la evaluación.
- Coloque los puntos principales al principio de cada sección y los menores después. Abra cada párrafo, enunciando el punto que trata.
- Argumente las conclusiones y recomendaciones con evidencias.
- Coloque la información técnica, incluso la matriz de diseño y los instrumentos de encuestas, en un apéndice.
- Designe tiempo para ¡¡¡revisar, revisar y revisar!!!
- Consiga que alguien que no conozca el material lo revise y corrija el texto del borrador y pídale que identifique lo que no se incluyó o lo que no esté claro.
- Si es posible, busque a un revisor externo con experiencia en los temas y conocimiento de metodología de evaluaciones, para que revise la propuesta final y sugiera cambios al documento cuando sea necesario. Si no es factible la revisión de pares, solicite a un colega no asociado con la evaluación que revise el informe.

Resumen ejecutivo

En el *resumen ejecutivo* de un informe se identifican las preguntas de la evaluación, se describe la metodología utilizada y se resumen los hallazgos, conclusiones y recomendaciones del informe. Proporciona una forma como el lector puede captar rápidamente los mensajes principales del informe. El resumen ejecutivo no es simplemente una versión condensada de la sección de conclusiones del informe. No es un rompecabezas que promete revelar información posteriormente. Debe servir como un documento independiente para los lectores demasiado ocupados para leer el informe principal.

Resumen ejecutivo: Resumen corto de un informe en donde se identifican las preguntas de la evaluación, se describe la metodología utilizada y se resumen los hallazgos, conclusiones y recomendaciones del informe.

De acuerdo con Scriven (2007, p. 1), el fin del resumen ejecutivo es resumir los resultados, no únicamente el proceso:

> A través de todo el proceso de evaluación, manténgase preguntándose cómo será el resumen general, basado en lo que haya aprendido hasta el momento, y cómo se relaciona con las necesidades del cliente, las partes interesadas y la audiencia; esto le ayuda a enfocarse en lo que todavía falta por hacer para saber qué es lo que más importa.

El resumen ejecutivo debe ser corto, lo mejor son dos páginas y más de cuatro, demasiado. Vale la pena repetir este punto: ¡dos páginas a espacio sencillo son lo mejor y más de cuatro páginas a espacio sencillo son demasiado!

Un resumen ejecutivo debe incluir los componentes siguientes:

- Un breve panorama general o un párrafo introductivo que exprese el propósito del estudio y el tema de interés, escrito de forma que capte la atención del lector.
- Una descripción de la evaluación con las preguntas principales, más un breve enunciado sobre el alcance y la metodología de la evaluación.
- Suficiente información de antecedentes para situar el estudio en contexto.
- Un resumen de los hallazgos principales del informe. Utilizar el juicio para establecer qué hallazgos son los más importantes para la audiencia.
- Una forma de dirigir los lectores a las páginas de información en el texto.
- Las principales conclusiones y recomendaciones.

Cuerpo del informe

El cuerpo del informe de una evaluación debe contener los elementos siguientes, divididos usualmente en capítulos (en un informe más corto, secciones):

- Introducción.
- Descripción de la evaluación.
- Hallazgos.
- Conclusiones.
- Recomendaciones.

La introducción del informe incluye los componentes siguientes:

- Propósito de la evaluación.
- Información de antecedentes.
- Metas y objetivos del programa, descritos mediante un modelo de teoría del cambio.
- Preguntas de la evaluación.

La descripción breve de la evaluación incluye los elementos siguientes:

- Propósito.
- Alcance.
- Preguntas.
- Metodología.
- Limitaciones.
- Participantes (junta asesora, empresa consultora).

Los hallazgos siguen a la descripción de la evaluación. Al escribir esta sección, los evaluadores deberían:

- Presentar los hallazgos de forma que la audiencia pueda comprenderlos claramente.
- Incluir sólo los más importantes.
- Organizarlos alrededor de las preguntas del estudio y los temas o los problemas principales.
- Utilizar diagramas, tablas y otros elementos gráficos para resaltar puntos importantes.

Las últimas partes del informe son las conclusiones y las recomendaciones, que a menudo los lectores leen primero. Los evaluadores tienen con mucha frecuencia dificultades para distinguir los hallazgos de las conclusiones. Los hallazgos describen lo que se encontró en la evaluación y pueden relacionarse con el cumplimiento o el incumplimiento de algún criterio. Deben estar soportados por la evidencia.

Las conclusiones se basan en la evaluación profesional de los hallazgos, y deben referirse a cada objetivo secundario de la evaluación, y también al objetivo general del proyecto, el programa o la política. No debe presentarse nueva información en la sección de conclusiones.

Las recomendaciones llaman a la acción e indican lo que en el informe se desea que hagan el cliente o los otros interesados. A menudo son difíciles de proponer. No deben ser demasiado preceptivas, para reducir la prerrogativa de la administración de identificar soluciones específicas a los problemas identificados. Por ejemplo, en una evaluación se podría recomendar desarrollar una política de precios para actividades de asistencia técnica. En ella no se propondría la política o se especificarían en detalle los requerimientos de la política. Sin embargo, podrían especificarse componentes básicos que deberían incluirse en la política de precios.

Al mismo tiempo, las recomendaciones no pueden ser tan generales que no especifiquen. Deben ser claras y lo suficientemente específicas para que todos comprendan lo que es necesario hacer para seguirlas, qué organización o unidad debe actuar y cuándo debería ejecutarse cada una.

En los informes no deben incluirse "listas de lavandería" de recomendaciones. Los evaluadores deben limitar el número de recomendaciones principales a unas tres o cuatro. Es mejor agrupar las recomendaciones de manera que se encuentren en número manejable (tres o cuatro) con partes secundarias, según se necesite. Debe tenerse en cuenta el tono de las recomendaciones. Es importante recordar que los informes no toman las decisiones y que éstas las toman las personas.

Las recomendaciones no sirven de mucho si no se actúa sobre ellas. Una forma de hacer seguimiento a ellas es establecer un sistema de seguimiento que permita a las partes interesadas seguir la implementación de las recomendaciones de la evaluación. Se hace seguimiento a cada recomendación de una evaluación y al progreso hecho para implementarla, incluyendo:

- La fecha de la recomendación.
- Quién es responsable de actuar.
- Respuesta/progreso.

001El camino.indd 445 08/11/2010 03:27:59 p.m.

En el cuadro 13.2 se presenta una matriz sencilla que puede utilizarse para un sistema de seguimiento de recomendaciones. En esta matriz, los evaluadores llenan las dos primeras columnas, pero los administradores hacen el seguimiento y se aseguran de que el designado está poniendo en práctica la recomendación.

La Corporación Financiera Internacional (CFI), el brazo del sector privado del Banco Mundial, ofrece otro ejemplo. La CFI tiene su propio sistema de seguimiento de las recomendaciones. Los informes de evaluación preparados por el Independent Evaluation Group (IEG) de la CFI incluyen recomendaciones para la Administración de la CFI y la Management Response de la misma. Estas recomendaciones se discuten en el IFC Board's Committee on Development Effectiveness (CODE), que espera informes periódicos de estado sobre las recomendaciones, incluyendo su nivel y estado de adopción. El IEG, con la administración de la CFI, desarrolló el Management Action Tracking Record (MATR).

El MATR está diseñado para mantener la integridad del proceso de presentación de informes, en el cual ni el IEG ni la CFI pueden modificar las clasificaciones finales. En el gráfico 13.1 se ilustran las dos etapas del MATR. En la primera, el IEG y la CFI acuerdan los indicadores para evaluar la implementación de cada nueva recomendación. En la segunda, el estado y nivel de adopción de cada recomendación activa se actualizan periódicamente y se informan al CODE. No es necesario que las clasificaciones del IEG y la CFI sean las mismas. Cuando las recomendaciones se implementan, se reemplazan o ya no son pertinentes, se desactivan. A las recomendaciones del IEG no aceptadas por la administración de la CFI no se les hace seguimiento.

Presentar información en forma visual

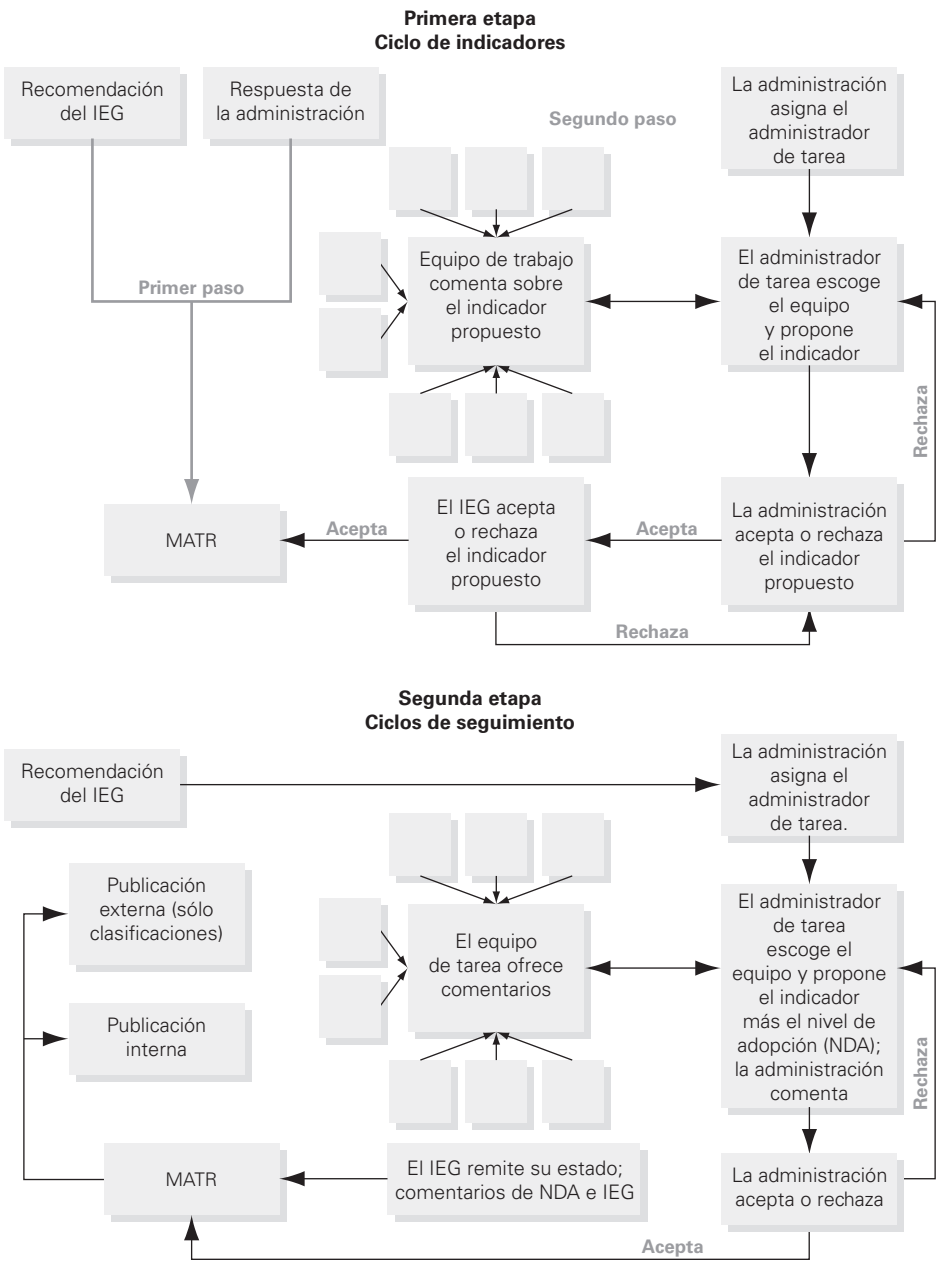
La información visual puede hacer que un informe sea interesante, comunicar información en forma más clara que el texto y atraer la mirada de los lectores hacia puntos particulares. Por consiguiente, el uso eficaz de gráficos puede aumentar el impacto de un informe de evaluación.

Cuadro 13.2 *Sistema de seguimiento de recomendaciones*

Recomendación	*Fuente*	*Fecha*	*Responsable*	*Respuesta / Progreso*
1.				
2.				
3.				
4.				

Fuente: los autores.

Gráfico 13.1 *El MATR de dos etapas de la CFI para el seguimiento de las recomendaciones*



Fuente: Independent Evaluation Group of the Internacional Finance Corporation, 2008.

Ilustraciones

Las ilustraciones ayudan a clarificar puntos del texto. Sin embargo, siempre deben utilizarse con alguna razón y nunca sólo por decoración. Las ilustraciones sobre la cubierta del informe pueden relacionarse con el tema general; las del interior del informe deben tener una razón específica y concreta para estar allí.

Todas las ilustraciones deben nombrarse en el texto, en el cual se debe indicar lo que debe ver el lector en ella. Ejemplos de ilustraciones que pueden utilizarse en una evaluación incluyen mapas, bosquejos, croquis y fotografías. Cuando se necesite, debe obtenerse el permiso para utilizar las ilustraciones.

Mapas

Los mapas pueden utilizarse en los informes para:

- Indicar la ubicación de un programa.
- Presentar un contexto.
- Indicar el alcance geográfico o difusión de un programa.
- Mostrar la base para tomar una muestra.
- Indicar tasas o niveles en la topografía de una zona, utilizando patrones o isogramas (Cummings, 2003).

Para aprovechar lo mejor de los mapas es necesario:

- Asegurarse de que sean fáciles de leer y comprender (los patrones y el sombreado deben distinguirse fácilmente).
- Utilizar la última versión.
- Identificar la fuente.
- Incluir una escala, cuando sea apropiado.

Bosquejos y croquis

Como parte de una recopilación de datos los evaluadores algunas veces piden a los participantes hacer bosquejos (ilustración 13.1), que pueden utilizarse para:

- Agregar interés a un informe.
- Personalizarlo.
- Reflejar un enfoque metodológico.
- Remplazar ilustraciones más sofisticadas si no se cuenta con la capacidad técnica para crearlas.
- Introducir humor.
- Permitir que se vean directamente las impresiones de los participantes.

001El camino.indd 448 08/11/2010 03:27:59 p.m.

Los croquis pueden utilizarse para ilustrar cómo funciona algo o cómo se relaciona un objeto con otro. Son muy útiles cuando es más difícil describir un proceso o una situación que dibujarlo.

Fotografías

Las fotografías son adiciones útiles a los informes. En ellas, se puede:

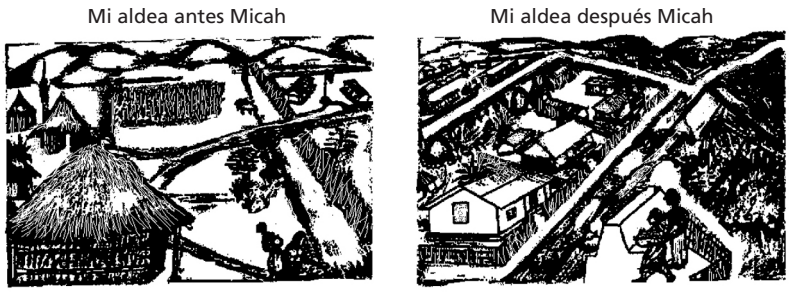
- Presentar un contexto.
- Indicar el grado del trabajo de campo (progreso).
- Capturar observaciones directas (por ejemplo, tipos de casas, condiciones de hacinamiento en vecindades).
- Familiarizar a la audiencia con la situación de campo.
- Presentar evidencia (Cummings, 2003).

A semejanza de otros tipos de ilustraciones, las fotografías nunca deben utilizarse sólo para decoración. Levin, Anglin y Carney (1987) revisaron y resumieron información sobre el uso de imágenes y de ilustraciones en los materiales. Concluyeron dos asuntos sobre los efectos que tienen las imágenes en el proceso de aprender en prosa:

- Cuando las ilustraciones son pertinentes al contenido, pueden esperarse ganancias desde moderadas hasta sustanciales en el aprendizaje.
- Cuando las ilustraciones NO son pertinentes al contenido o, lo que es peor, están en conflicto con él, NO pueden esperarse ganancias en el aprendizaje y quizá podría esperarse incluso confusión.

Así, debe haber una razón específica y concreta para incluir las fotografías y otras imágenes utilizadas en el informe.

Ilustración 13.1 *Impresión de un niño de la aldea antes y después de una intervención*



Fuente: Cummings, 2003.

Diagramas y gráficos

Diagrama organizativo: Diagrama que ilustra la jerarquía y la estructura de responsabilidad dentro de una organización.

Los cuadros sinópticos y los gráficos ofrecen representaciones visuales de los datos. Idealmente, cuentan su propia historia, sin la necesidad de una narración.

Un *diagrama organizativo* ilustra la jerarquía dentro de una organización (gráfico 13.2). Tales diagramas identifican con claridad y brevedad las responsabilidades y estructuras de mando.

Los diagramas organizativos se incluyen con mucha frecuencia en los informes de evaluación, porque el primer paso para comprender la estructura de una entidad es conocerla. La mayor parte de los programas de procesamiento de texto tienen una característica que permite crear dichos diagramas.

Diagrama de Gantt

Los *diagramas de Gantt* se utilizan a menudo para la planeación (gráfico 13.3). Son particularmente útiles para la administración de un proyecto, en especial su planeación.

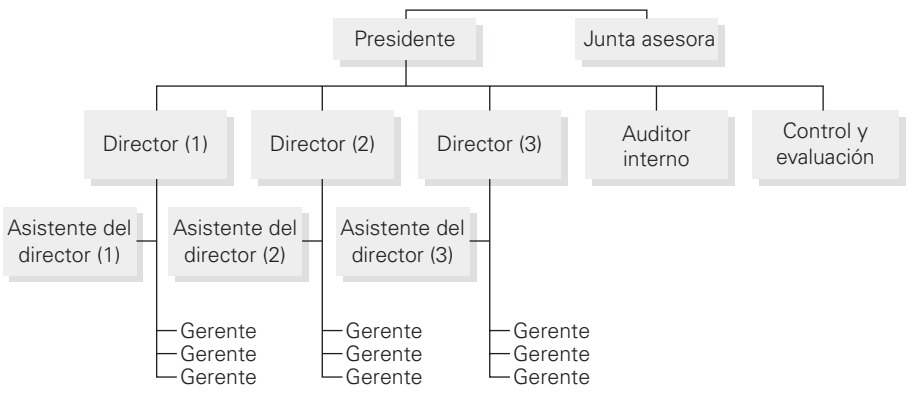
Gráficos

Un *gráfico* debe transmitir un mensaje claro con poco escrutinio.

Cada gráfico debe tener un título, un número y una línea fuente. El título del gráfico debe indicar el año de los datos (si se aplica) (cuadro 13.3).

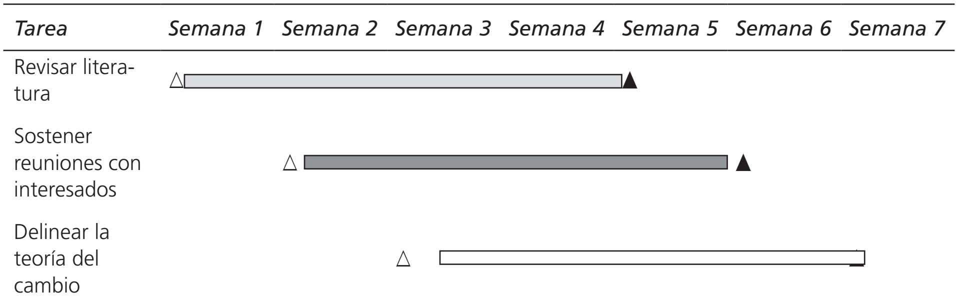
Por lo menos, cuatro tipos de gráficos son útiles para representar datos, cada uno de ellos sirve para un propósito distinto (cuadro 13.4).

Gráfico 13.2 *Ejemplo de diagrama organizativo*



Fuente: los autores.

Gráfico 13.3 *Ejemplo de diagrama de Gantt*



Fuente: los autores.

Cuadro 13.3 *Partes de los gráficos*

Nombre de la parte	*Descripción*
Título	Todos los gráficos y diagramas deben tener *título*, de modo que la audiencia conozca el mensaje del gráfico de inmediato.
Eje horizontal o eje x	El *eje horizontal* o *eje x* es la línea horizontal de un gráfico de línea o de barras, que representa una variable (por ejemplo, el tiempo).
Eje vertical o eje y	El *eje vertical* o *eje y* es la línea vertical de un gráfico de línea o de barras, que representa una segunda variable (por ejemplo, costos).
Origen	El *origen* es el punto en donde se encuentran los ejes vertical y horizontal.
Líneas de cuadrícula	Muchos diagramas incluyen líneas de cuadrícula como ayuda para comparar los datos, mostrando claramente niveles. Sólo debe utilizarse un número limitado de cuadrículas para evitar una apariencia confusa.
Títulos de los ejes	Los *títulos del eje x* y *del eje y* son muy importantes, e identifican lo que se mide y las unidades de medición (años, metros, libras, millas cuadradas, toneladas cúbicas, dólares, grados). Por ejemplo: • Costos (en US$). • Distancia (en km).
Escalas de los ejes	El eje *x* y el eje *y* requieren *escalas* apropiadas para mostrar los valores. Debe escogerse cuidadosamente la escala para incluir el rango completo de valores de los datos. Deben escogerse las proporciones entre las escalas de los ejes a fin de ilustrar mejor la relación entre las variables.
Valores actuales	Muchos gráficos y diagramas incluyen también los *valores actuales* de las entradas, presentadas como texto adicional dentro del gráfico. Estas adiciones son útiles para que el lector capte la situación real.
Coordenada	Una *coordenada* es el punto de un gráfico en el cual se encuentra el valor *x* con el valor *y.* La forma como esto se represente (un punto, un pico, la parte superior de una barra) dependerá del tipo de gráfico que se use.

Fuente: los autores.

001El camino.indd 451 08/11/2010 03:28:00 p.m.

Diagrama de Gantt: Diagrama utilizado para ilustrar la programación de un proyecto.

Gráfico: Representación visual de la relación entre dos o más piezas de datos.

Gráfico de línea: Gráfico que se elabora conectando una serie de puntos de datos con una línea; se utiliza usualmente para mostrar cambios a través del tiempo.

Gráfico de línea. Un *gráfico de línea* describe usualmente cómo cambia una variable (o variables) a través del tiempo (gráficos 13.4 y 13.5). Por ejemplo, los evaluadores podrían utilizar un gráfico de línea para mostrar cómo suben o bajan los costos de los alimentos a través de los meses, los cambios en la población durante muchos años, o las notas de los estudiantes cada día en un período de seis semanas.

Los gráficos de línea pueden mostrar un elemento o varios cuando cambian en un mismo período de tiempo y son una buena forma de mostrar datos continuos, esto es, datos de intervalos o razones. Los datos de intervalos son aquellos que se dividen en rangos y en los cuales la distancia entre los intervalos es significativa. Ejemplos de datos de intervalos son las cuentas de ingreso, los años de educación, o el número de votos. Los datos de razones son datos de intervalos que tienen también un verdadero punto cero. El ingreso es una razón porque cero dólares es realmente "no ingreso".

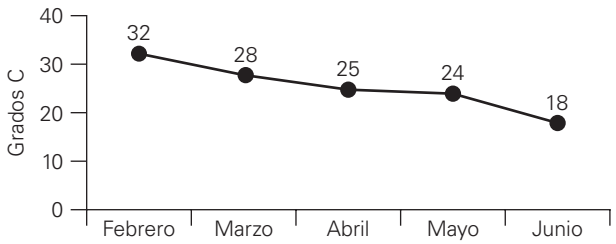
Gráfico de barras. En los *gráficos de barras* se utilizan barras (rectángulos alargados) para representar cantidades y permitir comparar números. Deben titularse cuidadosamente para asegurar que el lector comprenda qué factores se representan.

Hay dos clases de gráficos de barras: sencillos, que muestran información sobre una sola variable, y múltiples, que dan información para más de una variable. Las barras pueden ser de forma vertical u horizontal.

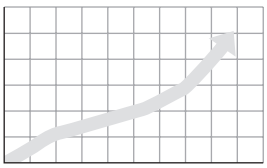
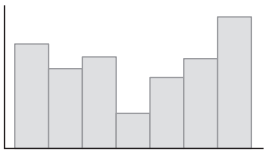
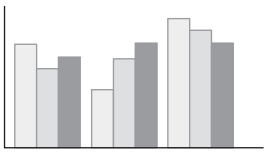
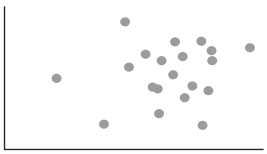
Gráfico de barras: Tipo de gráfico en el cual se utilizan rectángulos horizontales o verticales para representar y comparar cantidades.

Para comparar datos de dos o más grupos con la misma variable se utiliza un gráfico de barras múltiples. Por ejemplo, en una evaluación se desea comparar la tasa de recuperación de minas terrestres en tres regiones distintas de un país. Se podría utilizar un gráfico de barras múltiples para representar la información. Otro ejemplo de uso de gráficos de barras dobles sería el de comparar las respuestas de jóvenes, hombres y mujeres a un cuestionario.

Gráfico 13.4 Reducciones en la temperatura promedio en la provincia A, febrero-junio, 2008

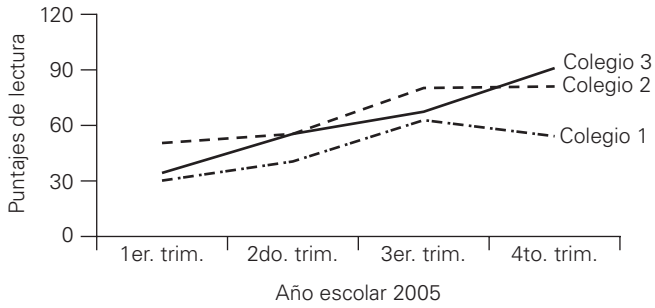


Cuadro 13.4 *Tipo y propósito de los distintos tipos de gráficos*

Tipo de gráfico	*Ejemplo*	*Propósito*
De línea		Mostrar tendencias en el tiempo.
De barras sencillo		Comparar características lineales o de una dimensión.
De barras múltiple		Comparar dos o más características con los valores de una variable común.
De pastel		Mostrar partes de un total.
Diagrama de dispersión		Mostrar tendencias o relaciones.

Fuente: los autores.

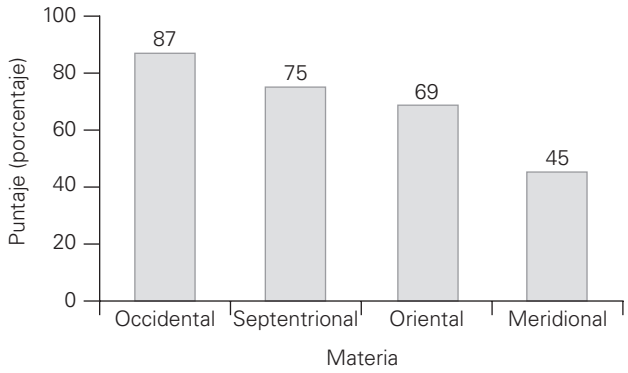
Gráfico 13.5 *Los puntajes de lectura difieren en tres colegios del distrito, 2004-2005*



Los gráficos de barras se utilizan con mucha frecuencia para presentar datos nominales o categóricos, que no tienen un orden, y la asignación de números a las categorías es puramente arbitraria (1 = Este, 2 = Norte, 3 = Sur). Estas categorías deben explicarse con claridad en las leyendas.

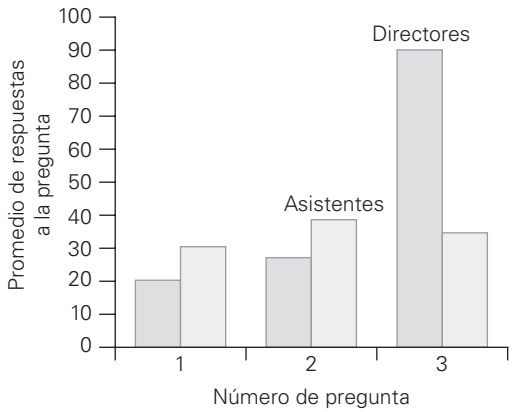
Los siguientes son ejemplos de gráficos de barras. En el gráfico 3.16 se muestra el puntaje obtenido en una prueba de cuidado de niños efectuada a cuatro mujeres embarazadas. Un *gráfico de barras sencillas* muestra los datos de una variable particular para un solo grupo (gráfico 13.6). Un *gráfico de barras múltiples* presenta datos de una variable particular para más de un grupo (gráfico 13.7).

Gráfico 13.6 *Los puntajes en la prueba sobre cuidado de niños varían según la región*



Fuente: los autores; datos fabricados.

Gráfico 13.7 *Las respuestas a las preguntas son diferentes para asistentes y directores*



Fuente: los autores; datos ficticios.

Gráfico de pastel. Un *gráfico de pastel* puede ser útil cuando el número de secciones es pequeño (gráfico 13.8), y no debe utilizarse cuando el número de secciones supera unas ocho o es menor que tres.

Gráfico de pastel: Gráfico circular dividido en secciones en el cual cada sección muestra la magnitud relativa del elemento descrito.

Diagrama de dispersión. Un *diagrama de dispersión* es un gráfico que representa la relación de un conjunto de datos sobre los ejes vertical y horizontal (gráfico 13.9). Si las variables están correlacionadas, será evidente una línea recta o curva. Cuanto mejor la correlación, más cercanos estarán los datos a la línea recta o curva. Un diagrama de dispersión que no presente ningún patrón particular indica que no existe una relación clara entre las dos variables.

Tablas

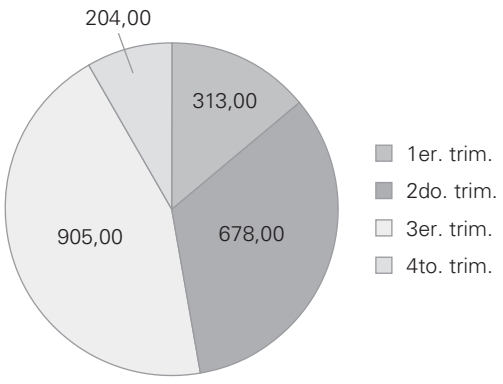
Las tablas se utilizan para presentar la información en forma organizada. Hay dos tipos de tablas que pueden utilizarse en un informe: tablas de datos y tablas de clasificación (matrices).

Tablas de datos

Puede ser útil presentar información numérica en tablas, llamadas tablas de datos, solamente si los conjuntos de datos son pequeños (Tufte, 1983). Las tablas para datos mayores generalmente constituyen la base para presentar los datos en otras formas, por ejemplo, gráficos de línea y de barras, y se colocan usualmente en los anexos.

Al igual que con las imágenes y las ilustraciones, la audiencia no sabrá automáticamente qué buscar en una tabla. Es útil que el título de una tabla describa lo que deben

Gráfico 13.8 *El tercer trimestre presenta los mayores costos de electricidad en dólares, 2008*



Fuente: los autores; datos ficticios.

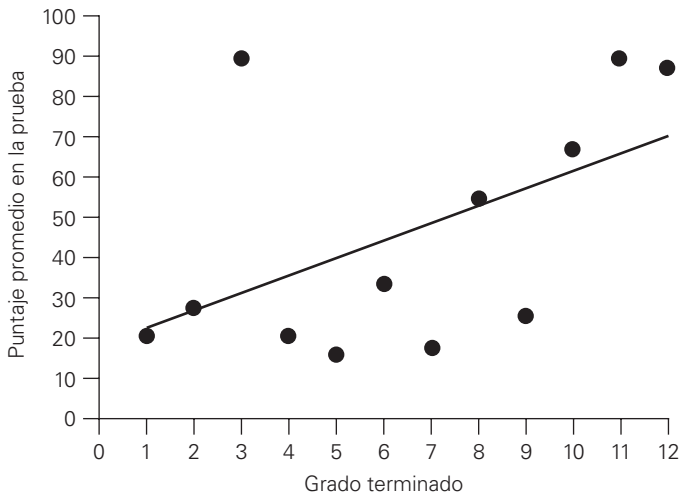
ver y cómo pueden relacionar la información. Una corta descripción de qué debe verse debería incluirse también en la narración del informe. Al presentar datos en una tabla deben incluirse las fuentes de los datos y el año en que se recopilaron.

Eherenburg (1977) resume los principios que guían el diseño de tablas para presentar información:

- Redondear los números a no más de dos dígitos significativos. Esto ayuda a la audiencia a hacer las comparaciones. (Nota: sugerimos redondear a enteros; la audiencia rara vez desea el detalle de puntos decimales.)
- Presentar sumas y promedios de filas y columnas (según sea apropiado) para ayudar a la audiencia a hacer comparaciones de las celdas individuales.
- Colocar los datos más importantes en columnas, ya que esto le permite al lector hacer comparaciones con facilidad.

Al decidir el formato de la tabla, tener presente que demasiadas líneas (divisorias de las celdas de la tabla) dificultan la lectura. Esto se muestra en las dos tablas siguientes, en las que el cuadro 13.5 da un ejemplo de datos en una tabla con muchas líneas. El cuadro 13.6 muestra los mismos datos en una tabla con menos líneas. Nótese cómo los datos son el foco de la segunda tabla y no las líneas. También nótese que la última fila de la tabla muestra los promedios de las columnas con datos.

Gráfico 13.9 *Comparación de puntajes de prueba y nivel de grado con puntajes promedio*



Fuente: los autores; datos ficticios.

Cuadro 13.5 *Ejemplo de datos en una tabla con muchas líneas*

Información demográfica sobre los participantes

Número del participante	*Altura*	*Peso*	*Edad*	*Distrito*
1	44	30	7,2	Norte
2	46	35	7,1	Este
3	40	20	7,6	Norte
4	32	22	7,2	Sur
5	29	23	7,0	Sur
6	50	38	7,8	Norte
7	44	30	7,3	Oeste
8	44	28	7,3	Oeste
9	42	30	7,5	Este
10	48	45	7,9	Sur
Promedio	38,09	27,36	6,72	

Cuadro 13.6 *Ejemplo de datos en una tabla con pocas líneas*

Información demográfica sobre los participantes

Número del participante	*Altura*	*Peso*	*Edad*	*Distrito*
1	44	30	7,2	Norte
2	46	35	7,1	Este
3	40	20	7,6	Norte
4	32	22	7,2	Sur
5	29	23	7,0	Sur
6	50	38	7,8	Norte
7	44	30	7,3	Oeste
8	44	28	7,3	Oeste
9	42	30	7,5	Este
10	48	45	7,9	Sur
Promedio	38	27	7,0	

Fuente: los autores (datos fabricados).
Nota: N = 10.

Tablas de clasificación (matrices)

Una tabla de clasificación, o matriz, tiene un diseño que muestra cómo se ha organizado una lista de cosas de acuerdo con un factor diferente (cuadro 13.7). Al menos dos factores de clasificación indican similitud o diferencia entre las cosas que se clasifican.

Cuadro 13.7 *Ejemplo de tabla de clasificación*

Estrategias de reducción de la pobreza: países del estudio de caso

País	*Fecha de inicio*	*Años de implementación*	*Revisión completa*
Etiopía	17 Sept. 2002	4,7	Sí
Guinea	25 Julio 2002	4,9	Sí
Mauritania	6 Feb. 2001	6,3	Sí
Mozambique	25 Sept. 2001	5,7	Sí
Tanzania	30 Nov. 2000	6,3	Sí

Fuente: adaptado de Banco Mundial, 2004.
Nota: datos de mayo de 2007.

Las tablas de clasificación ayudan también a ilustrar información compleja. Una matriz de diseño es una tabla de clasificación.

Mejorar la eficacia de la información visual

De acuerdo con Edward Tufte, experto en la presentación de información visual, las presentaciones gráficas deben lograr lo siguiente:

- Mostrar los datos.
- Inducir al observador a pensar en la sustancia del gráfico y no en la metodología, diseño gráfico, tecnología de producción gráfica o cualquier otra cosa.
- Evitar distorsionar lo que se intenta decir con los datos.
- Representar muchos números en un pequeño espacio.
- Explicar los grandes conjuntos de datos.
- Estimular la vista para comparar diferentes piezas de datos.
- Revelar los datos con distintos niveles de detalle, desde una visión amplia hasta la estructura fina.
- Integrarse estrechamente con las descripciones estadística y verbal de los datos (Tufte, 1983).

Tufte se refiere a la cantidad de tinta utilizada para presentar los datos mismos como "tinta de datos". La mayor parte de la tinta utilizada para crear un gráfico debería ser tinta de datos, y debe utilizarse el mínimo posible de tinta en las cuadrículas, rótulos u otros elementos de distracción (recuadro 13.2). Él se refiere a la decoración que no agrega información adicional como "basura de diagrama".

En el gráfico 13.10 se presenta un ejemplo de gráfico sobrecargado con "basura de diagrama". El sombreado y las cuadrículas no enriquecen en absoluto el mensaje del gráfico y los rótulos de los datos en la parte superior de cada columna distraen la atención del lector para observar los patrones. El gráfico 13.11 presenta la misma imagen sin basura. La mayor parte de la tinta de este gráfico es tinta de datos.

Recuadro 13.2 *Consejos para crear gráficos eficaces*

Los consejos siguientes pueden ayudar a maximizar la tinta de datos y a incrementar la eficacia de los gráficos:

- Presentar datos suficientes para comunicar el mensaje, pero limitar la cantidad de datos presentados en un solo gráfico.
- Asegurar que el gráfico comunique un mensaje y que éste tenga sentido para una persona razonable.
- Utilizar un tipo de letra legible, no menor de 10 puntos de tamaño. Evitar usar demasiados tamaños y diseños de letras. Facilitar la lectura a la audiencia (fijar el tipo horizontalmente, de ser posible, y evitar interferencia de otros elementos del gráfico).
- Concentrarse en los datos y no en los contenedores de los datos.
- Evitar cuadrículas pesadas, demasiadas cuadrículas ocupadas, más marcas de las necesarias, representación redundante de datos simples, recuadros, sombras, apuntadores y leyendas, cuando sea posible.
- Reemplazar las cajas alrededor de las imágenes trazadas a lo largo de una cuadrícula *x* o *y*.
- Ocultar las cuadrículas en los gráficos de barras.
- Utilizar marcas (sin líneas) para mostrar la ubicación actual de los datos *x* y *y*.
- Podar los gráficos, reemplazando las barras con líneas, eliminando tinta que no es de datos, eliminando las líneas de los ejes, e iniciando los ejes *x* y *y* en los valores de los datos (marcos de rango).
- Mantener las escalas honestas y especificarlas con claridad. Suministrar tanta información de escala como sea necesario, pero incluirla tan discretamente como sea posible.
- Evitar patrones concurridos e innecesarios.
- Tener cuidado con las sombras que pueden no mostrar la información deseada al imprimirse.
- Ser generoso en el uso de espacios en blanco, que proporcionan alivio visual.
- Indicar la fuente de la información.

Fuente: Tufte, 1983.

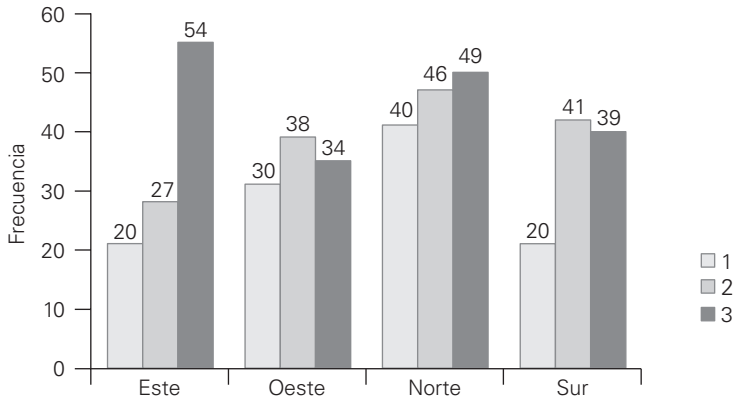
Hacer una presentación oral

Muchas personas tienen un gran temor de hablar en público. Este temor puede mitigarse preparándose bien y practicando la presentación con bastante anticipación.

Al planear una presentación considere las preguntas siguientes:

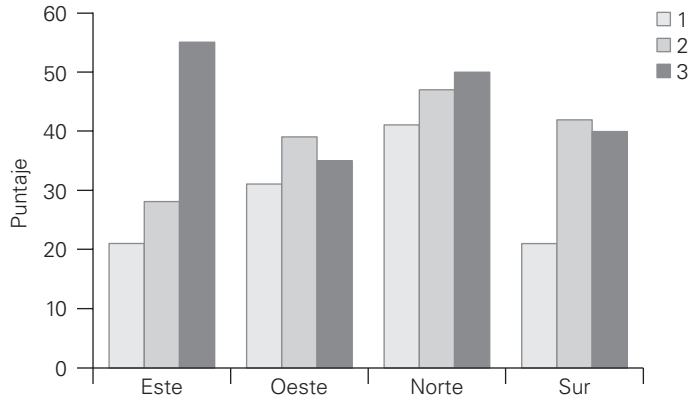
- ¿Quién es la audiencia? ¿Qué espera? ¿Cuánto detalle desea?
- ¿Cuál es el objeto de mi presentación? ¿Cuáles son los tres mensajes más importantes que quiero que recuerde la audiencia? ¿Qué quiero que haga la audiencia con la información que presento?
- ¿Existen problemas de lenguaje o técnicos para comunicar esta información a la audiencia?

Gráfico 13.10 *Ejemplo de gráfico innecesariamente confuso*



Fuente: los autores.

Gráfico 13.11 *Ejemplo de gráfico estilizado que maximiza la tinta de datos*



Fuente: los autores.

- ¿Cómo averiguo por anticipado cómo puede responder la audiencia a mi presentación?
- ¿En cuánto tiempo tengo que hacer la presentación?
- ¿Con qué recursos audiovisuales puedo contar (diapositivas, proyecciones en lo alto)?

Al preparar una presentación tenga presente la audiencia, concéntrese en los mensajes principales y respete la sencilla regla comprobada: "Dígales lo que les va a decir, dígaselo, y dígales lo que les dijo".

Una de las formas de mejorar la calidad de una presentación es practicar. Ensaye la presentación a solas antes de ensayarla frente a otras personas. Solicite retroalimentación después del ensayo y ajuste la presentación de acuerdo con ella. Asegúrese de que la presentación toma el tiempo previsto y no más.

Durante la presentación háblele a la audiencia, no a sus notas. Haga contacto visual con muchas personas de la audiencia. Si está utilizando proyector y pantalla, imprima una copia de las diapositivas y colóquelas frente a usted, de manera que no tenga que estar parado a espaldas de la audiencia y leyendo la pantalla.

Uso de ayudas visuales

Los elementos visuales enriquecen una presentación. Pueden consistir en diapositivas, proyecciones en lo alto, carteleras y volantes. (Si se utilizan diapositivas o proyecciones en lo alto, asegúrese de tener un plan de respaldo, en caso de que falle la electricidad o algún equipo.)

La información escrita refuerza la información transmitida oralmente. Por esta razón, especialmente al hacer una presentación a personas cuyo idioma nativo puede ser diferente, es a menudo una buena idea distribuir una pequeña cantidad de volantes al principio de la presentación (recuadro 13.3). Esto tiene también la ventaja de permitir que las personas agreguen sus anotaciones en la copia. Sin volantes, ¡no es raro ver a las personas tratar frenéticamente de escribir la presentación! La desventaja de distribuir volantes al principio de la presentación es que la gente tiende a adelantarse con los volantes en lugar de concentrarse en lo que dice el presentador. Si por esta razón se mantienen los volantes hasta el final, es importante avisar a la audiencia que se les repartirán. Si la presentación incluye datos o tablas complejas, considere la posibilidad de repartirlas al presentar el tema.

Recuadro 13.3 *Consejos para preparar presentaciones en PowerPoint y volantes*

Los volantes pueden mejorarse, teniendo presente estas indicaciones:

- Minimice la cantidad de palabras. Limite el texto a no más de ocho líneas por diapositiva o por proyección en lo alto.
- Incluya abundante espacios en blanco.
- Utilice materiales visuales claros.
- Mantenga el número de diapositivas por página en los volantes en dos para mejor legibilidad.

Uso de programas de presentación

Los programas de presentación, por ejemplo PowerPoint, le dan un aspecto más profesional a la presentación. Sin embargo, deben utilizarse juiciosamente, ya que pueden convertirse en un simple *show* de diapositivas. Para mantener interesante una presentación, el orador debe limitar el número de diapositivas, utilizándolas para transmitir solamente los puntos de mayor importancia (recuadro 13.4).

El uso excesivo de diapositivas puede hacer mortífera una presentación y se arriesga también "convertirlo todo en una promoción de ventas", y en "establecer el dominio del orador sobre la audiencia" (Tufte, 2003). Como anota Tufte:

> Las audiencias soportan en consecuencia una despiadada secuenciación, una maldita diapositiva tras otra. Cuando se apila la información en el tiempo, es difícil entender el contexto y evaluar las relaciones. El razonamiento visual funciona usualmente con mayor eficacia cuando la información pertinente se presenta a cada lado. Con mucha frecuencia, cuanto más intenso el detalle, mayores la claridad y la comprensión. Esto es especialmente válido para los datos estadísticos, en los que el acto analítico fundamental es hacer comparaciones (Tufte, 2003).

Resumen

Comunicar los resultados de una evaluación es de importancia tan crucial como realizar la evaluación: si los mensajes de una evaluación no se comunican, no hay mucha justificación para haberla realizado en primer lugar. Si una evaluación se presenta por escrito o en forma oral, debe adecuarse a las necesidades de la audiencia.

Tanto los informes escritos como los orales pueden mejorarse con el uso de gráficos. Respetando los principios delineados en este capítulo, los evaluadores pueden aumentar el impacto que tengan sus evaluaciones.

Recuadro 13.4 *Consejos para el uso de diapositivas*

Los consejos siguientes pueden mejorar la eficacia de las diapositivas en una presentación:

- Empiece con una diapositiva que capture la atención de la audiencia. Dé a la audiencia 10 segundos para leerla y luego emplee los cinco minutos siguientes hablando sobre por qué es importante el punto que se presenta en la diapositiva.
- Utilice diapositivas que ilustren puntos importantes, hechos sorprendentes, enunciados difíciles de captar oralmente y conclusiones importantes.
- Muestre la diapositiva y luego cuéntele a la audiencia la historia que hay detrás de ella.

Fuente: basado en Taylor, 2007.

Actividades del capítulo 13

Ejercicio de aplicación 13.1: mejorar el informe de la evaluación

Desarrolle una lista de criterios para evaluar qué tan bien transmite un informe de evaluación su metodología, hallazgos, conclusiones y recomendaciones a su audiencia prevista. Utilizando estos factores, evalúe un informe recientemente elaborado sobre su campo de acción. Déle una nota basado en su evaluación de cada criterio (A para excelente, B para muy bueno, C para adecuado y N para necesita mejorar). Identifique las mejoras que podrían hacerse al informe de modo que comunique en forma más eficaz sus mensajes.

Ejercicio de aplicación 13.2: adecuar un informe para una audiencia

Para el informe que revisó en el ejercicio anterior, identifique las distintas audiencias que podrían estar interesadas en sus hallazgos o metodología. Considere qué aspectos de la evaluación serían de mayor interés para cada audiencia, el nivel de detalle que sería de interés para cada audiencia y la mejor forma de comunicar los mensajes principales del informe para satisfacer las necesidades y preferencias de cada grupo. Utilizando su análisis, haga una lista de verificación para mostrar qué clase de comunicación debe recibir cada audiencia, cuándo y de quién.

Referencias y lecturas adicionales

Busuladzic, Dino y James Trevelyan. 1999. *Demining Research: An Ergonomic Aspect of Humanitarian Demining.* University of Western Australia, Perth. http://www.mech.uwa.edu.au/jpt/demining/tech/dino/ergonomics.html.

Cummings, Harry. 2003. "Using Graphics in Development Evaluations". Programa Internacional para la Capacitación en Evaluación para el Desarrollo (IPDET); presentación, Carleton University, Ottawa, junio 16-julio 11.

Druker, Phil 2006. *Advanced Technical Writing.* University of Idaho, Moscú, Idaho. http://www.class.uidaho.edu/adv_tech_wrt/week14/conclusion_recommendation_final_report.htm.

Ehrenberg, A. S. C. 1977. "Rudiments of Numeracy". *Journal of the Royal Statistical Society* 140: 277-97.

Independent Evaluation Group of the International Finance Corporation. 2008. *Enhancing Monitoring and Evaluation for Better Results: Biennial Report on Operations Evaluation in IFC 2008.* Washington D.C.: Banco Mundial. http: www.ifc.org/.../BROE2008_FullReport/$FILE/BROE+for+disclosurefinal+V8-final.pdf.

Kaplan, Craig S. 2003. "Presentation Tips". Craig Web Experience. http://www.cgl.uwaterloo.ca/~csk/presentations.html.

Lawrenz, Frances, Arlen Gullickson y Stacie Toal. 2007. "Dissemination: Handmaiden to Evaluation Use". *American Journal of Evaluation* 28 (3): 275-89.

Lentz, Michelle 2007. "Tufte and Visual Presentation". http://www.writetech.net/2007/02/presentation_st.html.

Lester, P. M. 2000. *Visual Communication: Images with Messages.* 2ª ed. Kentucky: Wadsworth Cengage Learning.

Levin, J. R., G. J. Anglin y R. N. Carney. 1987. "On Empirically Validating Functions of Pictures in Prose". En *The Psychology of Illustration,* Vol. 1, eds. D. A. Willows y H. A. Houghton, 51-85. Londres: Springer-Verlag.

MIGA. 2007. *Annual Report 2007-08.* http://www.miga.org/documents/IEG_MIGA_2007_Annual_Report.pdf.

Scriven, Michael. 2007. "Key Evaluation Checklist". http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf.

Taylor, Dave. 2007. "Use PowerPoint to Enhance Your Presentation, Not Cripple It". http://www.intuitive.com/blog/use_powerpoint_to_enhance_your_presentation_not_cripple_it.html.

Torres, R., H. S. Preskill y M. E. Piontek. 1997. *Evaluation Strategies for Communicating and Reporting.* Thousand Oaks, CA: Sage Publications.

Tufte, Edward R. 1983. The *Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

———. 1990. *Envisioning Information.* Cheshire, CT: Graphics Press.

———. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press.

———. 2003. "PowerPoint Is Evil: PowerPoint Corrupts. PowerPoint Corrupts Absolutely". *Wired* 11.09, septiembre. http://www.wired.com/wired/archive/11.09/ppt2.html.

———. 2006. *Beautiful Evidence.* Cheshire, CT: Graphics Press.

———. 2008. The *Work of Edward Tufte.* http://www.edwardtufte.com/tufte/.

Wallgren, Anders, Britt Wallgren, Rolf Persson, Ulf Jorner y Jan-Aage Haaland. 1996. *Graphing Statistics and Data.* Thousand Oaks, CA: Sage Publications.

World Bank Operations Evaluation Department. 2004. "The Poverty Reduction Strategy Initiative: An Independent Evaluation of the World Bank's Support through 2003". Banco Mundial, Washington, DC.

Sitios web

Statistics Canada. *Using Graphs.* http://www.statcan.ca/english/edu/power/ch9/using/using.htm.

Torok, George. *Presentation Skills Success.* http://www.presentationskills.ca/.

CAPÍTULO 14

Guiar al evaluador: ética, política, estándares y principios orientadores

Los evaluadores tienen muchas tareas, entre ellas la planeación, la organización y el diseño de las evaluaciones, y la recopilación, el análisis y la presentación de los datos. Deben tratar también con presiones internas y externas. Se les puede solicitar que hagan modificaciones al plan, la organización o el informe de la evaluación para satisfacer las necesidades de otras personas. Algunas veces, las modificaciones propuestas son aceptadas; otras, pueden suscitar consideraciones éticas o políticas.

La ética y la política constituyen problemas para todos los evaluadores, especialmente para aquellos que trabajan en países con gobernabilidad deficiente y con un historial de corrupción. Las presiones internas dentro de una organización para el desarrollo pueden crear también problemas éticos. En este capítulo se tratan los aspectos éticos y las consideraciones políticas de las evaluaciones.

Comportamiento ético

La ética es un conjunto de valores y creencias que orientan las elecciones. La ética trata grandes preguntas, como "¿Cómo deberíamos vivir?". Piénsese en esta pregunta.

001El camino.indd 465 08/11/2010 03:28:00 p.m.

¿Quién define "cómo" (método), "deberíamos" (ambición) ["nosotros"] (un grupo que busca consenso) y "vivir" (seres con cuerpos) (World IQ, 2008)? La ética depende de las personas que toman las decisiones y se basa en su opinión sobre lo que es correcto, equivocado, bueno y malo.

Las leyes se promulgan como intento de identificar y controlar el comportamiento ético, pero las leyes no pueden cubrir todas las situaciones posibles. Además, el comportamiento puede ser legal, y aun así no ético (por ejemplo, recibir un pequeño regalo de alguien que va a ser evaluado o cambiar el tono del informe para hacerlo más positivo o negativo, a pesar de la fuerza de la evidencia que lo soporta).

A menudo, los evaluadores se enfrentan a situaciones en las que no es claro lo que hay que hacer, y existen muchas zonas grises, pero siempre se espera que realicen las evaluaciones éticamente. Puesto que la base para decidir qué es ético se basa en valores y creencias, diferentes culturas establecen diversas leyes. Por esta razón, es importante ser consciente de lo que es legal y lo que se considera ético en el país donde se lleva a cabo la evaluación.

Los evaluadores deben también ser conscientes de las leyes que requieren informar sobre un crimen o la sospecha de un crimen. Por ejemplo, en los Estados Unidos, el acoso sexual es un crimen. Se define ampliamente como los avances sexuales mal recibidos, las solicitudes de favores sexuales y otra conducta verbal o física de naturaleza sexual, cuando el sometimiento a la conducta o su rechazo afecta implícita o explícitamente el empleo o el desempeño en el trabajo de una persona, o crea un entorno de intimidación, hostil u ofensivo (U.S. Equal Employment Opportunity Commission, 2002). El no reportar el acoso sexual sospechoso podría crear responsabilidad legal.

Los evaluadores deben ser conscientes de las leyes del país en donde trabajan y de la orientación y los estándares de su propia organización. Muchas organizaciones para el desarrollo han elaborado y emitido estándares o lineamientos éticos.

Corruptibilidad y falacias de la evaluación

Los evaluadores deben tener la capacidad de "decirle la verdad al poder". Una disposición mental independiente, inquisitiva y analítica es una de las cualidades necesarias para el evaluador. Fitzpatrick, Sanders y Worthen (2004) presentan cinco formas de "corruptibilidad de la evaluación":

- Disposición a tergiversar la verdad y producir hallazgos positivos, como resultado de conflicto de intereses u otros pagos o penalidades percibidos (dicha disposición puede ser consciente o inconsciente).
- Intrusión de opiniones no sustanciadas debida a prácticas de evaluación negligentes, caprichosas o no profesionales.
- "Hallazgos" de evaluación "sombreados" como resultado de intrusión de los prejuicios o nociones preconcebidas del evaluador.

- Incentivos a clientes o participantes para que cooperen haciendo promesas que no pueden cumplir.
- No cumplir los compromisos que podrían haberse cumplido.

House (1995) examina la corruptibilidad desde una perspectiva ligeramente distinta, e indica que los evaluadores pueden simplemente entender mal sus responsabilidades. Identifica cinco "falacias de evaluación":

- *Clientismo:* la falacia de que hacer cualquier cosa que el cliente solicite o que beneficie al cliente es éticamente correcto.
- *Contractualismo:* la falacia de que el evaluador debe seguir el contrato escrito sin cuestionamiento, aun si es perjudicial para el bien público.
- *Metodologismo:* la creencia de que seguir métodos de indagación aceptables asegura que el comportamiento del evaluador es ético, aun cuando algunas metodologías puedan en realidad multiplicar los dilemas éticos del evaluador.
- *Relativismo:* la falacia de que todos los datos de opinión que el evaluador recopile de varios participantes deben tener igual ponderación, como si no hubiese base para asignar apropiadamente a las opiniones de grupos periféricos menor prioridad que la asignada a grupos más esenciales.
- *Pluralismo/elitismo:* la falacia de permitir que se asigne mayor prioridad a los poderosos, porque el evaluador piensa que tienen más prestigio y potencia que los impotentes o sin voz.

Identificación de problemas éticos

Morris y Cohn (1993) encuestaron a los miembros de la American Evaluation Association acerca de su opinión sobre los problemas éticos. La lista siguiente de problemas éticos es una modificación de la de su encuesta:

- El cliente decide cuáles deben ser los hallazgos antes de que tenga lugar la evaluación, o planee utilizar los hallazgos en forma éticamente cuestionable.
- El cliente declara ciertas cuestiones de investigación como fuera de los límites de la evaluación a pesar de su pertinencia.
- El cliente modifica deliberadamente los hallazgos antes de su divulgación.
- El cliente presiona al evaluador para que altere la presentación de los hallazgos.
- El cliente suprime o ignora los hallazgos.
- El cliente presiona al evaluador para que viole la confidencialidad.
- El cliente hace mal uso no especificado de los hallazgos.
- Se excluye a partes interesadas legítimas del proceso de planeación.
- El evaluador descubre comportamiento ilegal, no ético o peligroso.
- El evaluador es renuente a presentar los hallazgos completos por razones no especificadas.

- El evaluador está inseguro sobre su capacidad de ser objetivo o justo al presentar los hallazgos.
- Aunque sin presión del cliente o las partes interesadas para violar la confidencialidad, el evaluador se preocupa de que informar ciertos hallazgos podría representar una tal violación.
- Los hallazgos se utilizan como evidencia contra alguien.

Además de estos problemas, existen cuestiones éticas asociadas con el diseño de una evaluación. Las cuestiones éticas pueden surgir en los diseños con asignación aleatoria, al decidir quién recibe la intervención y quién hará parte del grupo de control. Si, en el curso de un proyecto, una intervención demuestra ser promisoria, quienes no la reciben pueden demandar el acceso a ella. Garantizar dicho acceso modifica el diseño de la evaluación y la evaluación.

Para que las evaluaciones sean útiles para los directores, las organizaciones para el desarrollo, participantes y ciudadanos, deben ser honestas, objetivas y justas. Corresponde al evaluador asegurar que los datos se recopilen, analicen e informen honesta y justamente.

La influencia al principio de una evaluación puede ser sutil. Algunas veces hay presiones para evitar ciertas clases de preguntas de evaluación, o encaminar la evaluación hacia terrenos menos sensibles. Pueden no plantearse ciertos temas que podrían reflejarse negativamente en la organización o el programa. Los clientes pueden decir "Sabemos y ustedes saben que existe un problema que debe resolverse y hemos iniciado ya la acción correctiva, pero no es necesario que esto se haga público y perjudicar así el apoyo para el programa". Puede haber resistencia a encuestar al personal, los participantes del programa o los ciudadanos porque podrían revelarse cuestiones sensibles (negativas). En otras situaciones, puede excluirse a ciertas personas de las reuniones o las entrevistas, o limitarse las salidas al campo por "restricciones de tiempo". El evaluador debe esforzarse por suscitar cuestiones que se intenta eludir, evitar ser nombrado por votación y asegurar que se escuchen y tengan en cuenta todos los puntos de vista.

Los intentos de influir en una evaluación se presentan también a menudo durante el curso de ésta. Si bien la mayor parte de los evaluadores reconocerían rápidamente un soborno, no siempre reconocen formas sutiles de influencia. Los ofrecimientos de amistad, las comidas o las actividades recreativas pueden constituir un gesto de bienvenida para alguien que se encuentra lejos de casa, pero pueden también representar un intento de influir en la perspectiva del evaluador y a la larga en el informe.

Algunas veces, alguien le proporciona al evaluador indicaciones sobre corrupción o fraude. El evaluador no debe intentar resolver si esta información es un intento de distraer la atención hacia otros asuntos, un intento del informante de desquitarse de alguien, o si se trata de corrupción real o de fraude. Los evaluadores no se capacitaron para ser investigadores especiales, y no deberían, por tanto, investigar para establecer si los alegatos son ciertos. Más bien, deberían informar sobre el alegato a la autoridad apropiada para su investigación.

En la mayoría de las organizaciones para el desarrollo existen procedimientos para informar mala conducta sospechosa o fraude. Por ejemplo, en el Banco Mundial, el Department of Institutional Integrity investiga alegatos de mala conducta del personal, al igual que alegatos de fraude y corrupción en las operaciones del Banco. Dichos departamentos cuentan a menudo con líneas de atención gratuitas para informar sobre mala conducta, fraude o abuso.

Los miembros del equipo de evaluación deben estar familiarizados con las políticas y estándares de la organización para el manejo de los alegatos de fraude, abuso y mala conducta. Estos procedimientos pueden formar parte de un manual para el personal, un folleto especial o del contrato de empleo. Los evaluadores deben tener presente que están ligados a las políticas y estándares de su organización.

El lema "No hacer daño" ciertamente se aplica a la evaluación. Salvo cuando se sospecha e informa el fraude o el abuso para su investigación, las evaluaciones no deben perjudicar a los participantes. Esto significa que las personas participantes nunca deben ser identificadas o colocadas en situaciones amenazadoras.

Si se asegura la confidencialidad de los participantes y se divulgan los nombres posteriormente, se causa daño. Es esencial proteger la confidencialidad, pero puede haber situaciones en las que es difícil. Por ejemplo, supóngase que los evaluadores de un programa educativo reciben información de algunos entrevistados de que el director está gastando el dinero del programa en beneficio propio. ¿Qué deberían hacer los evaluadores? Con revelar estos hallazgos se arriesga exponer a quienes los informaron de manera confidencial. No revelar el incidente para investigaciones especiales lleva a una posible complicidad en el acto. Esta situación ilustra los beneficios de las líneas de atención anónimas. Puede ser también útil hablar con un supervisor o un gerente (no implicado en el alegato) sobre la situación y las opciones posibles.

Los evaluadores deberían tener especial cuidado cuando se les solicita responder a reporteros, administradores de programas o aun miembros de la junta, "extraoficialmente". Como mínimo, deberían clarificar qué significa "extraoficial" para la persona que les solicita información de esa forma. Los evaluadores deben sentirse seguros y satisfechos con la definición dada antes de proceder.

Política y evaluación

La evaluación se realiza siempre en un contexto político; el sólo hecho de que se realice una evaluación puede utilizarse para adelantar una agenda personal o institucional. Comúnmente se dice que el conocimiento es poder; la evaluación proporciona conocimiento que puede utilizarse como zanahoria o como garrote para forzar el comportamiento. Es una actividad que a menudo crea ganadores y perdedores.

La política puede socavar la integridad de una evaluación y determinar cómo y el grado al cual se utiliza ésta. Las evaluaciones positivas contribuyen a asegurar más fondos, ampliar los proyectos piloto a programas de escala completa, mejorar la reputación y la influencia, y construir la carrera de los participantes en la intervención.

Las evaluaciones que identifican problemas graves pueden mejorar las intervenciones y los resultados futuros, pero pueden dar como resultado menores presupuestos para los programas, o su cancelación y la pérdida de reputación, poder e influencia de los involucrados en la intervención.

Causas de la política en la evaluación

La evaluación es una forma de conocimiento organizativo. Las luchas de poder sobre definiciones de la realidad son, por eso, inherentes al proceso de evaluación.

Murray (2002) argumenta que la razón por la cual es inevitable la política en la evaluación es que existe mucho campo para la subjetividad, la cual lleva a diferencias en las personas incluidas en la evaluación. Los evaluadores acumulan percepciones de la realidad de las partes interesadas y de las personas evaluadas. Las percepciones pueden diferir, causando con frecuencia desacuerdos en distintas etapas de la evaluación y dando salida al comportamiento político.

Murray (2002, p. 2) atribuye la base para los desacuerdos a "problemas inherentes con elementos técnicos de métodos de evaluación y debilidades muy comunes en muchos seres humanos". También describe debilidades técnicas y humanas en evaluaciones que pueden tener efectos políticos. Cada tipo de debilidad se describe a continuación.

Debilidad técnica

La mayoría de las evaluaciones funcionan mejor cuando se miden contra metas, objetivos y estándares establecidos. Pero para los evaluadores, los clientes y otras partes interesadas puede ser difícil ponerse de acuerdo en cuanto a qué medir. Además, puede ser difícil determinar el foco de la evaluación. Una buena evaluación identifica la teoría del cambio y los supuestos subyacentes en el programa. Si se ha desarrollado con las partes interesadas del programa, la teoría del cambio puede asegurar un conocimiento común de las metas y objetivos del programa, las actividades y los resultados por los cuales debe ser responsable el programa. Así, una teoría del cambio puede contribuir a evitar conflictos potenciales sobre el conocimiento del programa, antes de que se conviertan en problemas políticos.

Murray (2002) identifica un segundo problema técnico común que ocasiona problemas políticos: evaluar un nivel de una organización y generalizar los hallazgos a otro. Esto causa problemas cuando los vínculos entre el desempeño de los individuos, los programas o las funciones y la organización como un todo no se establecen claramente. De nuevo, una teoría del cambio es una manera de identificar los supuestos subyacentes.

Debilidad humana

Los humanos hacen, con mucha frecuencia, lo que les conviene. Pueden tener prejuicios inconscientes que afectan la forma como realizan una evaluación o analizan sus resultados.

Cutt y Murray (2000) identifican tres factores humanos que pueden afectar una evaluación:

- La disposición mental de tipo "tenga buen aspecto-evite la vergüenza" (TBAEV).
- El fenómeno de la "interpretación subjetiva de la realidad" (ISR).
- Factores de confianza.

La disposición mental TBAEV identifica una característica humana: la gente desea tener éxito y evitar que se le asocie con el fracaso. La mayoría de las evaluaciones tienen la intención de revelar tanto los éxitos como los fracasos. Sin embargo, los evaluadores pueden omitir dar información sobre la medida completa de los fracasos por temor a que se les asocie con el fracaso. En una situación de TBAEV, la gente hará todo lo posible para explicar que los resultados negativos están por fuera de su control. De manera alternativa, pueden cuestionar el alcance y el enfoque de la evaluación (una razón para que los acuerdos por adelantado sobre el diseño de la evaluación sean tan importantes).

El fenómeno de la ISR aparece durante la interpretación de los datos de la evaluación. Dos personas que presencian el mismo evento pueden describirlo como "un maestro que pierde el control de la clase", o como "un maestro que se ha comprometido totalmente con la clase sobre un asunto". Los evaluadores tienen ideas y actitudes preconcebidas sobre lo que funciona. Los resultados de la evaluación pueden interpretarse de forma subjetiva con base en esas ideas y actitudes. Por ejemplo, el evaluador que piensa que un buen maestro es el que mantiene a los alumnos en su silla, hablando uno solo a la vez cuando se les solicita, tiene más probabilidad de ver los eventos como un caos o una pérdida de control. El evaluador cuyo hijo está en un salón de clases "abierto" tiene más probabilidad de ver el mismo evento como un ejemplo de un maestro que se compromete con los estudiantes.

El fenómeno de la ISR es una de las razones por las cuales este texto indica que los evaluadores deben hacer revisiones de la literatura para conocer lo que funciona bien, y lo que las evaluaciones temáticas han identificado en programas similares. Otra forma de reducir el sesgo de la realidad subjetiva es utilizar varios métodos de recopilación de datos, por ejemplo cuestionarios, entrevistas y observaciones estructuradas, con varias fuentes como una estrategia de triangulación.

Otro factor identificado por Cutt y Murray es el de la confianza. Puede activar (o causar) los factores TBAEV e ISR. La confianza es la creencia en la integridad o la capacidad de una persona. Si la gente considera que una persona carece de integridad o de capacidad, desconfía de ella. Pueden temer que esa persona pueda perjudicarla. La confianza se mide en grados, que varían de la confianza parcial (solamente en determinados contextos o sobre ciertos asuntos) hasta la confianza plena (en todas las situaciones). Cuando hay desconfianza, existe la probabilidad de que los fenómenos TBAEV e ISR ocasionen la entrada de la política en la relación. En los capítulos anteriores se han sugerido varias estrategias para generar confianza, desde incluir a las partes interesadas en el diseño de la evaluación hasta técnicas específicas de generación de confianza.

001El camino.indd 471 08/11/2010 03:28:01 p.m.

Identificación de juegos políticos

Es imposible mantener una evaluación completamente separada de la política. El reto está en manejar la situación política en la que se encuentran los evaluadores. Para ayudarlos Murray (2002) clasifica juegos de acuerdo con los papeles de las personas implicadas. La comprensión de estos juegos ayuda a los evaluadores a manejar la situación.

Juegos políticos de las personas que son evaluadas

Las personas que son evaluadas desean por lo general evitar un escrutinio formal de sus actividades, a lo cual pueden responder:

- Negando la necesidad de la evaluación.
- Reclamando que la evaluación les robaría mucho tiempo de su trabajo normal, o especialmente el más pesado y pidiendo que se posponga.
- Reclamando que la evaluación es una buena cosa, pero utilizando tácticas dilatorias.
- Buscando formar relaciones personales cercanas con los evaluadores, a fin de convencerlos de que pueden confiar en ellas.

Una vez iniciada la evaluación y la recopilación de los datos, surgen otros juegos políticos. Los proveedores de información pueden:

- Omitir o distorsionar la información que los evaluadores les solicitan, de modo que no se vean mal.
- Suministrar a los evaluadores enormes cantidades de información, dificultándoles separar lo pertinente de lo que no lo es.
- Salir con nuevos datos hacia el final de la evaluación.

Una vez se han recopilado los datos y están los evaluadores listos para interpretarlos, las personas que están siendo evaluadas pueden responder:

- Negando la existencia del problema.
- Disminuyendo la importancia del problema, arguyendo que ya lo conocían y están implementando cambios, o atribuyéndolo a otros, o a fuerzas que están por fuera de su control.
- Reclamando que la información ya no es pertinente porque las cosas han cambiado.

Juegos políticos de otras partes interesadas

Otras partes interesadas pueden también afectar la política de una evaluación. Diferentes partes interesadas pueden tener distintas agendas e intereses.

Si las partes interesadas no participaron en la identificación de las principales preguntas de la evaluación, pueden decidir que en la evaluación se examinaron cosas

equivocadas. Además, pueden intentar que otros, por ejemplo los medios, critiquen la organización e indiquen cómo debió haberse realizado la evaluación.

Juegos políticos de los evaluadores

Los evaluadores también hacen sus juegos en la evaluación. Durante la recopilación de los datos, algunos evaluadores trastocan el proceso recopilando su propia información "extraoficialmente". Esta información informal puede entonces entrar en la fase de interpretación de la evaluación.

La mayor parte de los juegos de los evaluadores ocurre durante la fase de interpretación de la evaluación. Los juegos incluyen:

- Desplazar o no expresar los estándares de medición.
- Aplicar criterios no expresados a la toma de decisiones.
- Aplicar valores no expresados y filtros ideológicos a la interpretación de los datos, por ejemplo, decidir que una fuente de datos no es confiable.
- Ignorar datos no consistentes con la conclusión del evaluador.

Manejo de la política en las evaluaciones

Puesto que la política es inevitable en las evaluaciones, es importante saber cómo manejarla. Durante el proceso de evaluación, el evaluador debe generar confianza. Idealmente, durante cada fase de una evaluación, se sostienen discusiones abiertas que les dan a todos los actores una oportunidad de discutir sus intereses, y por lo menos estar de acuerdo o en desacuerdo sobre sus diferencias. Murray (2002) discute la importancia de asegurarse de que todas las partes interesadas implicadas en la evaluación conozcan totalmente la lógica subyacente, y sugiere que la teoría del cambio (o modelo lógico) es una forma de articular la lógica de manera que haya poco campo para malos entendidos.

¿De qué otras formas se genera confianza? Habitualmente toma tiempo y muchos encuentros. Murray sugiere involucrar a todas las partes interesadas en el proceso de evaluación, en particular a quienes van a ser sometidos a evaluación.

Equilibrar a las partes interesadas con negociaciones

Para los evaluadores uno de los grandes retos es tener que tratar con muchas partes interesadas. Los evaluadores deben poseer fuertes destrezas de negociación para manejar los intereses y las agendas a menudo rivales de las muchas partes interesadas. Markiewicz (2005) presenta principios y prácticas útiles para negociar una evaluación.

Principios para negociar una evaluación

Los puntos siguientes fueron adaptados de la lista de Markiewicz (2005) de principios para negociar evaluaciones:

- Reconocer la naturaleza inherentemente política de la evaluación.
- Valorar la contribución de las distintas partes interesadas.
- Evaluar las posiciones de las partes interesadas y planear la evaluación.
- Asegurar que el evaluador sea un actor activo dentro de la comunidad de partes interesadas.
- Desarrollar las destrezas del evaluador como negociador que responde a un conflicto.
- Desarrollar destrezas en manejo de conflictos con distintas partes interesadas.

Una estrategia clave es organizar a las partes interesadas en grupos de referencia, comités de dirección o de asesoría para supervisar el proceso de la evaluación. Este enfoque es consistente con el modelo del "evaluador como facilitador" (no es usualmente aplicable a evaluaciones independientes cuyo propósito principal es evaluar la responsabilidad). Es importante que estos grupos tengan papeles y funciones definidos claramente. Los grupos o los comités deben fijar reglas de terreno que definan cuán activos deben ser los miembros en el proceso de evaluación.

Según Markiewicz (2005), una vez que el evaluador establece un nivel de credibilidad y aceptación con las partes interesadas, él o ella deben negociar áreas de conflicto o de disputa entre ellos. El evaluador debe actuar como catalizador que brinde ayuda a las partes interesadas para encontrar sus propias soluciones. Para esto debe contar con fuertes destrezas de comunicación, que incluyen la habilidad de escuchar de manera activa y reflexiva, hacer preguntas apropiadas y verificar el entendimiento. El evaluador debe también mantener centrado el proceso de negociación y alentar la interacción entre las partes interesadas.

Los evaluadores deben desarrollar destrezas de negociación. En algunos casos pueden arreglar la capacitación y la práctica adicional en destrezas de negociación (el evaluador como modelo facilitador). Otra forma como los evaluadores desarrollan destrezas de negociación es trabajar con pares para compartir experiencias (exitosas y fallidas) de resolución de conflictos.

Patton (1997) sugiere que por lo menos se hagan cuatro reuniones (más para proyectos de más largo plazo):

- Primera reunión: foco de la evaluación.
- Segunda reunión: métodos y herramientas de medición.
- Tercera reunión: instrumentación desarrollada antes de la recopilación de datos.
- Cuarta reunión: revisión de datos para llegar a un acuerdo sobre interpretaciones que llevan a los hallazgos.

Esto es similar a los pasos recomendados en este libro.

Markiewicz discute el papel activo que los evaluadores deben representar con las partes interesadas. Dos de las características que ella describe como valiosas son las de ser responsable y flexible, de modo que las partes interesadas puedan vincularse al

proceso. Discute también las dificultades que surgen si el evaluador se acerca mucho y tiene demasiadas relaciones interpersonales con las partes interesadas.

Patton sugiere permanecer concentrado en el proceso empírico y contribuir a que también lo hagan las partes interesadas, ya que esto ayuda a mantener las relaciones de manera objetiva, y evita la intrusión de sesgos o el mal uso de los hallazgos.

Práctica de negociar la evaluación

En la negociación de la evaluación, Markiewicz (2005) identifica tres etapas:

- Etapa inicial: se ponen al descubierto las posiciones.
- Etapa intermedia: tiene lugar la negociación activa
- Etapa final: se dan los pasos para lograr un consenso.

Para utilizar este modelo, el negociador de la evaluación debe tener varias destrezas de empatía y asertivas. Las destrezas de empatía generan un clima conducente a la negociación. Las asertivas le proporcionan estructura al proceso.

La empatía puede definirse como "el proceso de demostrar una comprensión precisa, deliberativa de las necesidades, intereses y posición de la otra parte" (Mnookin, Peppet y Tulumello, 2000, p. 46). Incluye dos componentes: la capacidad de ver el mundo con los ojos de la otra parte y la capacidad de expresar ese punto de vista en palabras. La empatía incluye traducir el conocimiento de la experiencia de la otra parte en una respuesta compartida.

Markiewicz (2005) piensa que la empatía es una característica importante para adquirir información acerca de otras metas, valores y prioridades. Se convierte en catalizador para inspirar la apertura en los demás y en herramienta persuasiva para la negociación.

Una vez que el evaluador tiene un buen conocimiento de los puntos de vista de cada interesado, él o ella deben parafrasear ese conocimiento para todas las partes interesadas presentes (Hale, 1998). El evaluador debe entonces preguntar a las partes interesadas si ese conocimiento es correcto y pedirles clarificar las diferencias que pueda haber. Una escucha activa y reflexiva ayuda al evaluador a comprender lo que se dice, hacer preguntas apropiadas y revisar la comprensión de lo que dicen las partes interesadas.

La resolución es la capacidad de expresar y defender las propias necesidades, los intereses y posiciones (Mnookin, Peppet y Tulumello, 1996). Cuando se trata de negociar evaluaciones, puede describirse como la autoridad facilitadora. Puede ser difícil equilibrar la empatía y la resolución. Mnookin, Peppet y Tulumello (1996) ven la empatía y la resolución como dos dimensiones interdependientes del comportamiento en la negociación. Cuando se utilizan juntas, pueden producir beneficios sustanciales en la negociación y un mejor conocimiento de las necesidades del interesado.

001El camino.indd 475 08/11/2010 03:28:01 p.m.

Estándares de evaluación y principios orientadores

Las asociaciones profesionales desarrollan estándares o lineamientos para ayudar a sus miembros a tomar decisiones éticas. Las asociaciones profesionales de América del Norte, Europa y Australia han establecido códigos de ética para los evaluadores.

La American Evaluation Association (AEA) desarrolló un conjunto de estándares y principios, plasmados en dos documentos, *Program Evaluation Standards* y *Guiding Principles for Evaluators*. Estos estándares y principios han servido con mucha frecuencia como plataforma para otros grupos, que los han adaptado y modificado a sus circunstancias o situaciones locales.

La AEA desarrolló los *Guiding Principles for Evaluators* para ofrecer orientación a los evaluadores en su práctica diaria. La mayor diferencia entre los estándares y los principios de la AEA es su propósito. Los estándares tratan sobre el desempeño profesional, mientras los principios de orientación tratan sobre los valores profesionales. Los estándares se centran en el producto de la evaluación, mientras los principios se centran en el comportamiento del evaluador. Ambos documentos ofrecen orientación sobre formas éticas y apropiadas de realizar evaluaciones.

Los Estándares de Evaluación de Programas

Los Estándares de Evaluación de Programas de la AEA se agrupan en cuatro categorías, a saber:

- Utilidad
- Factibilidad
- Corrección
- Precisión

La corrección incluye ocho estándares específicos:

- *Orientación al servicio:* atiende la necesidad de los evaluadores de servir no solamente los intereses de la agencia que patrocina la evaluación, sino también las necesidades de los participantes en el programa, la comunidad y la sociedad.
- *Acuerdos formales:* Incluye asuntos como seguir los protocolos, tener acceso a los datos, advertir claramente a los clientes sobre las limitaciones de la evaluación y no prometer demasiado.
- *Derechos de sujetos humanos:* incluye cosas como obtener un consentimiento informado, mantener el derecho a la privacidad y asegurar la confidencialidad.
- *Interacciones humanas:* es una extensión de los derechos de los sujetos humanos y hace que los evaluadores respeten la dignidad y el valor humano en todas las interacciones, y que ningún participante en la evaluación sea humillado o perjudicado.
- *Evaluación completa y justa:* aspira a asegurar que las fortalezas y debilidades de un programa queden descritas con precisión. El evaluador debe asegurar que no inclina el estudio para satisfacer al patrocinador o apaciguar a otros grupos.

- *Divulgación de los hallazgos:* se refiere a la obligación del evaluador de servir al público general que se beneficia tanto del programa como de su evaluación precisa, no solamente los clientes o patrocinadores. Los hallazgos deben divulgarse públicamente.
- *Conflicto de intereses:* se refiere a la necesidad de los evaluadores de hacer explícitos su predisposición y valores en forma tan abierta y honesta como sea posible, de modo que los clientes estén alerta a prejuicios que puedan insinuarse involuntariamente en el trabajo de incluso los evaluadores más honestos.
- *Responsabilidad fiscal:* incluye la necesidad de asegurar que todos los gastos sean apropiados, prudentes, y se encuentren bien documentados, inclusive el tiempo y el esfuerzo empleados en proveer, recopilar y facilitar la obtención de información solicitada por los evaluadores, y de explicar las evaluaciones a los distintos electorados.

La AEA sostiene que estos estándares de evaluación de programas son solamente estadounidenses, y pueden no ser apropiados para otros países sin adaptación. En el año 2000, la W.K. Kellog Foundation financió una reunión de evaluadores regionales y nacionales en Barbados, en la cual estuvieron representadas varias organizaciones internacionales de evaluación. Uno de los temas discutidos fue el de si –y cómo– los estándares de evaluación de la AEA se relacionan con otros países.

Uno de los resultados de la reunión fue la publicación de documentos ocasionales sobre este tema (Russon, 2000). En el primer documento ocasional, Taut (2000) sostiene que los Estándares de Evaluación de Programas desarrollados por la AEA son valores basados y que los valores difieren según las culturas. Ella investiga cómo los valores que tienen discrepancias afectan la funcionalidad de los estándares de la AEA fuera de los Estados Unidos.

Taut compara los valores de los estándares con las dimensiones de valor cultural en la literatura de distintas culturas. Las dimensiones de valor cultural que ella identifica como más importantes son:

- Individualismo frente a colectivismo.
- Jerarquía frente a igualitarismo (o distancia de poder).
- Conservadurismo frente a autonomía.
- Dominio frente a armonía.
- Evasión de incertidumbre.

Otras diferencias que ella nota entre las culturas son comunicación directa frente a indirecta, contexto alto frente a bajo y la importancia de la antigüedad.

Taut concluye que "lo que es útil y ético difiere en mayor medida según la cultura de lo factible y preciso… Se hace claro que las cuestiones sobre el Estándar de Corrección son muy dependientes de las influencias políticas y culturales" (2000, p. 24). Recomienda que los evaluadores describan sus sociedades con relación a las dimensiones culturales

tratadas en su documento, y recomienda también que consulten con colegas sus percepciones y con expertos para que orienten su análisis.

Los principios orientadores para evaluadores

La AEA se esfuerza por promover la práctica ética en la evaluación de programas, de personal y políticas. Con ese fin, desarrolló principios orientadores (1995) para asistir a los evaluadores en su práctica profesional, entre ellos están los siguientes:

- Investigación sistemática
- Competencia
- Integridad y honestidad
- Respeto por las personas
- Responsabilidad por el bienestar general y público.

El AEA Ethics Committee supervisó una revisión y una actualización importantes de los Principios en 2004. (Esta publicación y mayor información sobre los Principios Orientadores y los Estándares de Evaluación de Programas se encuentran en el sitio web de la AEA. Los vínculos para ambos documentos se ofrecen al final de este capítulo en la sección "Organizaciones de evaluación".)

Otras organizaciones de evaluación con propuestas de lineamientos o estándares, o definitivos, disponibles en Internet incluyen las siguientes:

- La African Evaluation Association (estándares y propuestas de lineamientos de evaluación).
- La Australasian Evaluation Society (lineamientos éticos).
- La Canadian Evaluation Society Guidelines for Ethical Conduct.
- La German Society for Evaluation (estándares).
- La Italian Evaluation Association (lineamientos).
- La Swiss Evaluation Society (estándares).
- La U.K. Evaluation Society (lineamientos de buenas prácticas).

Además de lineamientos para profesionales, en los lineamientos de buenas prácticas del Reino Unido se incluyen lineamientos para comisionados y participantes en la realización de evaluaciones.

Ética de evaluación para el sistema de la ONU

El Grupo de Evaluación de las Naciones Unidas (UNEG) aborda la ética en evaluación en sus *Ethical Guidelines* (UNEG, 2008). Estos lineamientos están basados en los "ideales profesionales mantenidos habitualmente y reconocidos internacionalmente" (UNEG, 2008, p. 3). Se aplican a la realización de evaluaciones por parte de su personal, consultores externos y evaluadores de organizaciones asociadas en todas las agencias de las Naciones Unidas.

Un perfil de los principios éticos de estos lineamientos es el siguiente:

- Intencionalidad de la evaluación que incluye:
  - Utilidad.
  - Necesidad.
- Obligaciones de los evaluadores:
  - Independencia.
  - Imparcialidad.
  - Credibilidad.
  - Conflictos de intereses.
  - Honestidad e integridad.
  - Responsabilidad.
- Obligaciones con los participantes:
  - Respeto por la dignidad y la diversidad.
  - Derechos.
  - Confidencialidad.
  - Prevención de perjuicios.
- Proceso de evaluación y producto:
  - Precisión, integridad y confiabilidad.
  - Transparencia.
  - Presentación de información.
  - Omisiones y maleficencia (UNEG, 2008).

La UNEG estableció también los Estándares de Evaluación en el sistema de la ONU. Sus estándares de ética incluyen lo siguiente:

- Los evaluadores deben ser sensibles a las creencias, maneras y costumbres, y actuar con integridad y honestidad en sus relaciones con todas las partes interesadas.
- Los evaluadores deben asegurar que sus contactos con individuos se caractericen por el mismo respeto con el que les gustaría ser tratados.
- Los evaluadores deben proteger el anonimato y confidencialidad de los informantes.
- Los evaluadores son responsables de su desempeño y sus productos.

Estándares del CAD

En el primer capítulo se discutió sobre la OECD Development Assistance Committee Network on Development Evaluation, el principal grupo para normas y estándares de evaluación para el desarrollo. Un resumen actualizado de las normas clave del CAD se encuentra disponible en http://www.oecd.org/dac/evaluationnetwork.

Para este capítulo son de particular aplicación los estándares propuestos de calidad de evaluación del CAD, que ofrecen guías sobre procesos y productos de la evaluación. Estos estándares estarán terminados entre 2009 y 2010, después de una fase de

prueba de tres años. La sección 6 de los lineamientos se centra en la independencia de los evaluadores y el grado al cual el equipo de evaluación puede trabajar libremente y sin interferencias. La sección 7 se dedica a la ética en la evaluación. Esta propuesta de guías indica que el proceso de evaluación, y de ahí el evaluador, "muestran sensibilidad al género, las creencias, las maneras y las costumbres de todas las partes interesadas y se realiza con integridad y honestidad. Se protegen los derechos y el bienestar de los participantes en la evaluación. Deben protegerse el anonimato y la confidencialidad de los informantes cuando se solicite o cuando lo requiera la ley" (p. 22).

Conflicto de intereses

El conflicto explícito o implícito de intereses es un problema importante que afecta potencialmente la credibilidad de los evaluadores y la responsabilidad de la evaluación. Para cada evaluación, los evaluadores deben indicar si están libres de conflicto de intereses o de apariencia de conflicto.

Algunas organizaciones han desarrollado lineamientos que rigen el conflicto de intereses. Por ejemplo, el Operations Evaluatign Group del Asian Development Bank les prohíbe al personal y a los consultores evaluar trabajos en los que ellos participaron. Ésta es una buena práctica para las unidades de evaluación.

Resumen

La ética es un conjunto de valores y creencias que orientan las elecciones. Los evaluadores pueden violar los estándares éticos realizando evaluaciones, a pesar de tener conflicto de intereses, ocuparse en prácticas de evaluación negligentes, o no profesionales, permitir que los perjuicios afecten su trabajo, hacer promesas que no pueden cumplir, o no cumplir sus compromisos. El malentendido en forma de clientismo, contractualismo, metodologismo, relativismo o pluralismo y elitismo, puede ocasionar también el comportamiento no ético.

La OECD Development Assistance Committee Network on Development Evaluation, el Grupo de Evaluación de las Naciones Unidas (UNEG), la AEA y otras agencias y asociaciones han desarrollado estándares, lineamientos y normas para definir y medir la calidad, la ética y las normas. Los evaluadores para el desarrollo deben familiarizarse con los estándares, lineamientos y normas pertinentes y modelar sus prácticas.

001El camino.indd 480 08/11/2010 03:28:01 p.m.

Actividad del capítulo 14

Ejercicio de aplicación 14.1: rosa y la evaluación agrícola

Rosa Márquez le cuenta a usted la historia siguiente y le pide su consejo. ¿Cuáles son los principales problemas éticos y cómo le aconsejaría a Rosa tratarlos?

"Rosa se reunió con funcionarios locales, funcionarios del programa y dueños de tierras para enterarlos de una próxima evaluación de un programa agrícola local. Con el transcurso de los años una comunidad ha recibido una cantidad sustancial de dinero para construir sistemas de irrigación, comprar fertilizantes, construir vías y adquirir equipos.

Ésta era la primera visita de Rosa a la zona. El miembro del equipo local, Eduardo, ya la había visitado en varias ocasiones y conocía a muchos de los propietarios. Él sugirió que salieran a comer y empezaran a armonizar.

Durante la cena, Rosa escuchó la conversación que sostenían Eduardo y los propietarios, quienes al parecer tenían una relación cercana con Eduardo, y le obsequiaron una caja de cigarros. Discutieron las necesidades de la zona. Los propietarios consideraban que necesitaban más recursos para utilizar con eficacia la tierra. Querían más equipo y reemplazar algunos de los agricultores, y también utilizar más fertilizante del permitido por las leyes ambientales. Eduardo estuvo de acuerdo con los propietarios y les dijo que la próxima evaluación podría ayudar, pues recomendaría que se hiciese una excepción con ellos.

La cena terminó con una invitación para que Rosa se uniese a uno de los propietarios para hacer una gira por la zona, seguida de un almuerzo con su familia. Rosa consideró que sería rudo no aceptar la invitación e hizo planes para reunirse con él al día siguiente. Ella habló brevemente con Eduardo después de la cena y le preguntó por qué estaba de acuerdo con el propietario. Éste le comentó que pensaba que los propietarios serían más colaboradores si sentían que obtendrían algo positivo de la evaluación.

Durante la gira del día siguiente, el propietario explicó lo duro que habían trabajado y el progreso que habían hecho contra todas las posibilidades. Le dijo a Rosa que contaba con que ella apoyaría sus esfuerzos, ya que si la evaluación era negativa, él y su familia no podrían sobrevivir. Como muestra de su aprecio, le regaló una gargantilla sobre la cual le dijo que había estado en su familia durante varias generaciones.

Después de la gira y el almuerzo con la familia del propietario, Rosa se reunió con el director del programa, quien había preparado un cronograma, indicando con quiénes se reuniría ella durante los tres días restantes. También preparó dos reuniones con la comunidad, que incluían a los propietarios, los trabajadores de extensión agrícola, los miembros de la comunidad de empresas vendedoras de equipos agrícolas y fertilizantes y exportadores de productos agrícolas. Al preguntar Rosa por qué no se incluía a los agricultores y sus familias, se le respondió que ellos no podían contribuir con nada que tuviese valor para la evaluación de la eficacia del proyecto. Ella preguntó si había otras personas en la comunidad con quienes debería hablar. Se le respondió que el director del programa había hecho todos los esfuerzos para asegurar que se incluyese a las personas adecuadas para facilitarle el trabajo de su evaluación del programa.

Referencias y lecturas adicionales

Cutt, James y Vic Murray. 2000. *Accountability and Effectiveness Evaluation in Nonprofit Organizations.* Londres: Routledge.

Fitzpatrick, Jody L., James R. Sanders y Blaine R. Worthen. 2004. *Program Evaluation: Alternative Approaches and Practical Guidelines.* Nueva York: Pearson Education Inc.

Hale, K. 1998. "The Language of Co-Operation: Negotiation Frames". *Mediation Quarterly* 16 (2): 147-62.

House, E. R. 1995. "Principles Evaluation: A Critique of the AEA Guiding Principles". En *Guiding Principles for Evaluators,* eds. R. Shadish, D. L. Newman, M. A. Scheirer y C. Wye, 27-34. New Directions for Program Evaluation N° 66. San Francisco: Jossey-Bass.

JCSEE (Joint Committee on Standards for Educational Evaluation). 1994. *The Program Evaluation Standards: How to Assess Evaluations of Educational.* Programs Thousand Oaks, CA: Sage Publications.

Markiewicz, Anne. 2005. "'A Balancing Act': Resolving Multiple Stakeholder Interests in Program Evaluation". *Evaluation Journal of Australasia* 4 (1-2): 13-21.

Molund, Stefan y Göran Schill. 2004. *Looking Back, Moving Forward: SIDA Evaluation Manual.* Estocolmo: Swedish International Development Cooperation Agency.

Mnookin, Robert H., Scott R. Peppet y Andrew S. Tulumello. 1996. "The Tension between Empathy and Assertiveness". *Negotiation Journal* 12 (3): 20-35.

———. 2000. *Beyond Winning: Negotiating to Create Values in Deals and Disputes.* Cambridge, MA: Harvard University Press.

Morris, M. y R. Cohn. 1993. "Program Evaluators and Ethical Challenges: A National Survey". *Evaluation Review* 17: 621-42.

Murray, Vic V. 2002. *Evaluation Games: The Political Dimension in Evaluation and Accountability Relationships.* http://www.vserp.ca/pub/CarletonEVALUATIONGAMES.pdf.

Patton, M. Q. 1997. *Utilization-Focused Evaluation: The New Century Text.* 3ª ed. Thousand Oaks, CA: Sage Publications.

Russon, Craig. 2000. The *Program Evaluation Standards in International Settings.* http://www.wmich.edu/evalctr/pubs/ops/ops17.pdf.

Stufflebeam, D. L. 1986. "Standards of Practice for Evaluators". Documento presentado en la reunión anual de la American Educational Research Association, San Francisco, abril.

Tassie, A. W., V. V. Murray, J. Cutt y D. Gragg. 1996. "Rationality or Politics: What Really Goes on When Funders Evaluate the Performance of Fundees?" *Nonprofit and Voluntary Sector Quarterly* 25 (3): 347-63.

Taut, Sandy. 2000. "Cross-Cultural Transferability of the Program Evaluation Standards". En *The Program Evaluation Standards in International Settings,* ed. Craig Russon. Serie de documentos ocasionales, Western Michigan University, Evaluation Center, Kalamazoo. http://www.wmich.edu/evalctr/pubs/ops/ops17.pdf.

UNEG (Grupo de Evaluación de las Naciones Unidas). 2005. *Norms for Evaluation in the UN System.* http://www.uneval.org/index.cfm?module=UNEG&Page=UNEGDocuments&LibraryID=96.

———. 2008. *Ethical Guidelines.* http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=102.

U.S. Equal Employment Opportunity Commission. 2002. http://www.eeoc.gov/facts/fs-sex.html.

World IQ. 2008. "Simple View of Ethics and Morals". http://www.wordiq.com/definition/Simple_view_of_ethics_and_morals.

Sitios web

Banco Interamericano de Desarrollo. 2004. *Proposal for Sectorwide Approaches (SWAps).* http://enet.iadb.org/idbdocswebservices/idbdocsInternet/IADBPublicDoc.aspx?docnum=509733.
Banco Mundial. *The World Bank Participation Sourcebook.* Washington, DC: Banco Mundial. http://www.worldbank.org/wbi/sourcebook/sbhome.htm.
Department for International Development. http://www.keysheets.org/red_7_swaps_rev.pdf.
Evaluation Center, Western Michigan University. http://www.wmich.edu/evalctr/Human Rights Education. http://www.hrea.org/pubs/EvaluationGuide/.
Institute of Development Studies. 1998. *Participatory Monitoring and Evaluation: Learning from Change,* IDS Policy Briefing 12 (noviembre). http://www.ids.ac.uk/ids/bookshop/briefs/PB12.pdf.
MEASURE Evaluation Project. *Monitoring and Evaluation of Population and Health Programs.* University of North Carolina, Chapel Hill. http://www.cpc.unc.edu/measure.
National Aeronautics and Space Act of 1958. http://www.hq.nasa.gov/office/pao/History/spaceact.html.
Treasury Board of Canada. *Linkages between Audit and Evaluation in Canadian Federal Developments.* http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/TB_h4/evaluation03_e.asp.
United Nations Population Fund. *UNFPA List of Evaluation Reports and Findings.* http://www.unfpa.org/publications/index.cfm.

Organizaciones de evaluación

African Evaluation Association. http://www.geocities.com/afreval/.
American Evaluation Association. http://www.eval.org.
Australasian Evaluation Society. http://www.aes.asn.au/.
Canadian Evaluation Society. http://www.evaluationcanada.ca.
European Evaluation Society. http://www.europeanevaluation.org.
German Society for Evaluation. http://www.degeval.de/.
Gobierno y ONG. http://www.eval.org/Resources/govt_orgs_&_ngos.htm.
Institute of Internal Auditors. http://www.theiia.org.
International Organization of Supreme Audit Institutions. http://www.gao.gov/cghome/parwi/img4.html.
Italian Evaluation Association. http://www.valutazioneitaliana.it/aiv/news.php.
Swiss Evaluation Society. http://www.seval.ch/en/.
United Kingdom Evaluation Society. http://www.evaluation.org.uk/.

Normas y principios orientadores

American Evaluation Association. 2004. *Guiding Principles.* http://www.eval.org/Publications/GuidingPrinciples.asp. (Versión abreviada en folleto: http://www.eval.org.)
———. 2005. *Program Standards.* http://www.eval.org/EvaluationDocuments/progeval.html.
Australasian Evaluation Society. *Ethical Guidelines for Evaluators.* http://www.aes.asn.au/about/Documents%20-%20ongoing/guidelines_for_the_ethical_conduct_of_evaluations.pdf.
Banco Asiático de Desarrollo. *Guidelines.* Operations Evaluation Group. http://www.adb.org/documents/guidelines/evaluation/independent-evaluation.pdf.
Canadian Evaluation Society. *Guidelines for Ethical Conduct.* http://www.evaluationcanada.ca.
German Society for Evaluation. *Standards.* http://www.degeval.de/standards/standards.htm.

Italian Evaluation Association. *Guidelines.* http://www.valutazioneitaliana.it/statuto.htm#Linee).
OCDE. *Principles for Evaluation of Development Assistance.* Development Assistance Committee. http://www.oecd.org/dataoecd/31/12/2755284.pdf.
———. *DAC Criteria for Evaluating Development Assistance.* Development Assistance Committee. http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html.
———. DAC Network on Development Evaluation. *Evaluating Development Co-Operation: Summary of Key Norms and Standards.* http://www.oecd.org/dac/evaluationnetwork.
United Kingdom Evaluation Standards. 2008. http://www.mfcr.cz/cps/rde/xbcr/mfcr/EvalStandards_UK.pdf and http://www.evaluation.org.uk/resources/guidelines.aspx.

CAPÍTULO 15

Mirando hacia el futuro

En los 14 capítulos anteriores se describieron "los caminos para obtener resultados" y en ellos se discutieron los fundamentos de la evaluación para el desarrollo; cómo preparar, diseñar y realizar una evaluación y cómo actuar profesionalmente. Esos capítulos nos traen hacia el presente. En este capítulo se revisa en forma breve dónde hemos estado y dónde estamos hoy antes de mirar a lo que el futuro depara a la evaluación para el desarrollo.

Del pasado al presente

Los cambios en el desarrollo se reflejan pronto en la evaluación para el desarrollo. En el año 2000, la reunión de líderes mundiales para aprobar la Declaración de Desarrollo del Milenio, y definir el desarrollo como una asociación global para la reducción de la pobreza volvió la atención a la medición de los resultados. La evaluación para el desarrollo respondió, pero todavía hay un largo camino por delante.

Este libro está dedicado a incrementar el uso del conocimiento evaluativo para la toma de decisiones. Para determinar la eficacia del desarrollo existen siete pasos críticos con los que puede mejorar la utilidad de la evaluación:

001El camino.indd 485 08/11/2010 03:28:01 p.m.

1. *Cambio a la evaluación basada en resultados*. Se le presta mayor atención a lograr resultados más allá de las medidas de producto, pero todavía no constituye una rutina en los trabajos para el desarrollo ni en la evaluación para el desarrollo. Sin saber cómo debería ser el éxito de una intervención en el corto, mediano y largo plazo, las organizaciones continuarán recompensando la provisión de bienes y servicios, sin estar seguras de que causen cambios en el comportamiento y los resultados en el mejoramiento institucional sostenible deseado.
2. *Identificar las teorías del cambio subyacentes en las intervenciones para el desarrollo*. La primera pregunta esencial para tratar el impacto no es "¿Cuál fue el impacto de la intervención?", sino más bien "¿Se necesitaba realmente la intervención y por qué se esperaría que ésta tuviese el impacto que tuvo?" Esta clase de análisis debe volverse una práctica estándar. Las teorías del cambio necesitan describir gráficamente la lógica del vínculo entre una intervención y sus resultados previstos, los supuestos clave que se hicieron sobre cómo y por qué funcionará, y los demás eventos mayores competidores en el ambiente a los que puedan atribuirse los cambios observados.
3. *Promover el uso apropiado de contrafactuales en los diseños de evaluación*. Los contrafactuales a menudo son necesarios para establecer el vínculo entre una intervención y sus efectos directos e impactos. Pero aunque los diseños experimentales (y cuasiexperimentales) constituyen diseños fuertes, no son un "patrón oro". La evaluación para el desarrollo debe centrarse en las preguntas que necesitan respuesta y seleccionar diseños apropiados a las preguntas. Puede ser más útil, por ejemplo, conocer el impacto marginal de una mayor duración de un programa, o la variación de una intervención en la implementación, como resultado de factores contextuales o características de los participantes que las diferencias en el impacto entre los grupos de tratamiento y de control. Pese a que el conocimiento de los diseños experimentales y cuasiexperimentales debe formar parte de cualquier capacitación sobre evaluación de impacto, dicha capacitación tiene que incluir también diseños no experimentales, por ejemplo los casos de estudio.
4. *Utilizar la cartografía, la contribución y la adicionalidad a nivel de países en lugar de buscar la atribución*. La atribución de resultados para el desarrollo a una sola entidad de desarrollo es difícil a nivel de proyectos; por lo general, es infructífera en evaluaciones más complejas, por ejemplo las realizadas a nivel de sector o de país. Con el reconocimiento del desarrollo como una asociación global, se requiere un cambio en los paradigmas. Un primer paso en la evaluación para el desarrollo es mapear las actividades de los principales donantes, prestamistas y ONG en un país determinado. Este análisis se haría preferiblemente en el país mismo, y se utilizaría luego como base para evaluar la coordinación de la ayuda y determinar la contribución de una organización para el desarrollo individual por medio de la adicionalidad.
5. *Reducir los costos e incrementar los beneficios de la evaluación a nivel de proyectos*. A pesar de la mucha retórica, la mayor parte de los recursos de una evaluación se

condicionan todavía a la evaluación de proyectos. Debería hacerse más uso de las evaluaciones grupales de proyectos relacionados de uno a tres años después del último desembolso de la financiación. En estas evaluaciones se podrían buscar patrones de efectos directos y poner a prueba si existen características o variables contextuales que sean críticas para el éxito.

6. *Generar capacidad de evaluación*. El mayor énfasis en la medición de los resultados ha llevado por sí mismo a una mayor demanda de especialistas en evaluación para el desarrollo. La Declaración de París y el Acuerdo de Agra han contribuido a centrar la atención en las necesidades de generación de capacidad de los países en desarrollo, y los donantes están respondiendo con mayor financiación, pero generar capacidad de evaluación será un esfuerzo a largo plazo. Los donantes necesitan también atender el problema de la generación de demanda de evaluadores en los ministerios y otras agencias en los países en desarrollo. Este lado de la ecuación sobre generar demanda no puede ignorarse.

   Los bancos de desarrollo multilateral, bilateral y otras organizaciones de ayuda deben atender la generación de capacidad de evaluación de su propio personal. Su personal dedicado a la evaluación, generalmente tomado de las filas internas de funcionarios de proyectos de desarrollo, carece con frecuencia de antecedentes sólidos de evaluación. Algunas veces son economistas o investigadores con otros antecedentes que, si bien están capacitados en métodos de investigación de ciencias sociales generalmente aplicables y econometría, carecen de capacitación en métodos cualitativos.

   La capacitación y otros esfuerzos de generación de capacidad en evaluación necesitan su propia evaluación para asegurar la calidad de los programas. Cuando la demanda supera la oferta, los programas de baja calidad pueden constituir un problema. Las exigencias de pericia en la capacitación en evaluación requieren verificación.

7. *Utilizar el conocimiento evaluativo*. Ninguno de los pasos listados hasta el momento representa mucha diferencia, a menos que se utilicen los hallazgos de la evaluación sólida para dar información sobre el desarrollo de la intervención siguiente. La evaluación no solamente tiene como propósito la rendición de cuentas, sino también el aprendizaje. Esto ha sido durante mucho tiempo el mantra de los evaluadores. Se conoce mucho sobre cómo incrementar la probabilidad de utilizar la evaluación en el contexto de la evaluación para el desarrollo. Ya ha llegado el tiempo de hacer énfasis en convertir este conocimiento en acción (Rist y Stame, 2006).

El futuro

Ninguno de nosotros cuenta con una bola de cristal para predecir el futuro con precisión. Sin embargo, podemos desarrollar visiones informadas de las tendencias y los desafíos futuros para la evaluación para el desarrollo. Cerramos este libro con algunas consideraciones sobre este tema.

Primero y lo más importante, vemos el aumento continuo del seguimiento y la evaluación (SyE) en casi cualquier país en desarrollo. Un desarrollo importante es efectuar viajes de estudio a países en desarrollo para conocer más sus sistemas de SyE. Éstos deben hacerlos personas de países en desarrollo y desarrollados, y pensamos que esfuerzos similares habrá no solamente en los ministerios y dependencias oficiales, sino también en los parlamentos, fundaciones, iglesias, asociaciones, otras organizaciones voluntarias y el sector privado.

La demanda de rendición de cuentas por los resultados no está desapareciendo. Su amenaza principal es la muerte por los sistemas sobrecargados de indicadores de seguimiento a los que se alimenta con demasiados datos, con el resultado de que el nivel de esfuerzo se hace insostenible.

Evaluación interna y externa

La evaluación para el desarrollo resistió el debate de métodos cuantitativos frente a cualitativos que estuvo en boga en la American Evaluation Association (AEA) en los años noventa. En realidad, el uso de métodos mixtos ha sido habitual en la comunidad para el desarrollo.

Lo más problemático en la evaluación para el desarrollo ha sido el asunto de si realizar evaluaciones externas o internas y en dónde, en la organización, ubicar las unidades de evaluación. Mientras los sistemas de seguimiento y el personal de SyE se consideran claramente como internos, en la evaluación se ha titubeado más. Esta cuestión es particularmente importante para quienes empiezan desde cero y no de un marco heredado.

¿Será la tendencia en la evaluación para el desarrollo a ser interna y una función combinada para quienes administran sistemas de seguimiento basados en resultados? ¿O tornará más hacia la estrategia de la Comisión Europea de favorecer la contratación de la evaluación a individuos y entidades externas?

Nuestro pronóstico es que la fuerte demanda continua por resultados significará que la evaluación para el desarrollo florecerá no solamente dentro de las unidades de SyE de la administración, sino también en solicitudes de evaluación a consultores y empresas externos. Así mismo prevemos mayores solicitudes y necesidades de evaluación independiente.

La evaluación realizada externamente no es necesariamente evaluación independiente. Si la evaluación se realiza externamente, pero es financiada por quienes administran el programa y está bajo su supervisión, es una evaluación interna y no debe considerarse independiente. Las evaluaciones independientes, las unidades de evaluación independiente y los evaluadores independientes deben cumplir principios de independencia, tales como los establecidos por el Evaluation Cooperation Group de los bancos de desarrollo multilateral (*ver* el capítulo 1), al margen de la reputación o la ética de la entidad que realiza la evaluación.

Una función de evaluación independiente puede estar ubicada dentro de una organización para el desarrollo (un ejemplo es el Independent Evaluation Group del Grupo

del Banco Mundial). Puede hacer parte de un parlamento o de otra organización legislativa. Pero dondequiera que esté ubicada, la unidad de evaluación independiente deberá evaluar formalmente su cumplimiento de los principios de independencia, utilizando una escala de clasificación como la presentada en el capítulo 1.

Desarrollo de capacidad de evaluación

La mayor demanda de SyE debe ir pareja con la generación de capacidad de evaluación. No es probable que una función de evaluación que dependa de unos pocos expertos satisfaga necesidades a largo plazo. Por lo general, debe desarrollarse la capacidad de evaluación a todos los niveles dentro de un país. Pronosticamos un aumento continuo en la generación de capacidad de evaluación en los países en desarrollo.

El desarrollo de la capacidad de evaluación incluye varias acciones para construir y fortalecer los sistemas de SyE en los países en desarrollo. Se centra en los niveles nacional y sectorial, e incluye muchos conceptos y herramientas relacionados, entre ellos, capacidades para calificar la eficacia del desarrollo, especificación de objetivos de proyectos, programas y cadenas de resultados, información sobre desempeño (incluso recopilación básica de datos), SyE de programas y proyectos, encuestas de evaluación de beneficiarios, revisiones sectoriales y auditoría del desempeño (Kusek y Rist, 2004; Mackay, 2007).

El desarrollo de capacidad de evaluación ayuda a los países a determinar en dónde empezar a generar la capacidad de evaluación (recuadro 15.1). Se han desarrollado varios instrumentos de diagnóstico (*ver*, por ejemplo, Banco Mundial, 1999).

El desarrollo de la capacidad de evaluación asegura que los hallazgos de la evaluación se encuentren disponibles para ayudar a los países en cuatro áreas clave:

- Asignar los recursos del gobierno para planeación, toma de decisiones y fijación de prioridades, sobre todo en el proceso del presupuesto.
- Conocer el funcionamiento de las actividades en curso a nivel de sector, programa y proyecto, de manera que puedan darse el aprendizaje y el mejoramiento.
- Responsabilizar a los administradores por el funcionamiento de las actividades que administran de modo que pueda responsabilizarse al gobierno por el desempeño. La noción de responsabilidad comprende el reconocimiento de que la gobernabilidad económica y un sector público responsable son centrales para la competitividad económica del país. Los mercados recompensan a los países capaces de manejar y filtrar el gasto público; la evaluación ofrece una herramienta para ayudar a los gobiernos a hacer esto.
- Demostrar hasta qué punto han sido exitosas las actividades para el desarrollo. Esto está demostrando ser cada vez más importante para los países que desean atraer recursos externos, en particular, dadas las presiones sobre las agencias internacionales de ayuda para el desarrollo para canalizar su ayuda a países en donde esfuerzos anteriores han tenido éxito. Además, el creciente énfasis de las

Recuadro 15.1 *Temas básicos para diagnosticar los sistemas de SyE de un gobierno*

1. Génesis del sistema existente de SyE: papel de los defensores o líderes del SyE, eventos clave que crearon la prioridad para la información de SyE (por ejemplo, elección de un gobierno orientado a la reforma, crisis fiscal).
2. El ministerio o la agencia responsable de administrar el sistema de SyE y planeación de las evaluaciones. Papeles y responsabilidades de las partes interesadas principales del sistema de SyE (los ministerios de finanzas, planeación, la presidencia, los ministerios sectoriales, la legislatura). Incentivos para que las partes interesadas tomen en serio el SyE (fuerza de la demanda por información de SyE). Posible existencia de varios sistemas no coordinados de SyE a niveles nacional y sectorial. Importancia de los problemas federales, estatales y locales para el sistema de SyE.
3. El entorno del sector público y si éste facilita o dificulta a los administradores un desempeño con altos estándares y responsabilidad en su desempeño. Existencia de reformas al sector público, por ejemplo, una estrategia de reducción de la pobreza, presupuesto orientado al desempeño, al fortalecimiento de capacidades de análisis de políticas, a la creación de una cultura del desempeño en el servicio civil, mejoras en la prestación de servicios (por ejemplo, estándares de servicio al cliente), descentralización gubernamental, mayor participación de la sociedad civil, o estrategia anticorrupción, que puedan beneficiarse con un mayor interés en la medición del desempeño del gobierno.
4. Los aspectos principales de la administración del sector público que apoye fuertemente el sistema de SyE, por ejemplo, la toma de decisiones presupuestaria, la administración del programa de planeación nacional o sectorial y las relaciones de responsabilidad (ante el Ministerio de Finanzas, la Presidencia, el Parlamento, los ministerios sectoriales, la sociedad civil).
5. El papel de la información sobre SyE en las distintas etapas del proceso presupuestario (asesoría y planeación política, toma de decisiones presupuestarias, revisión e información sobre el desempeño) y la posible desconexión entre el trabajo de SyE de los ministerios sectoriales y el uso de dicha información en el proceso del presupuesto. La existencia de cualquier desconexión entre el proceso del presupuesto y la planeación nacional. Oportunidades para fortalecer el papel del SyE en el presupuesto.
6. El grado al cual la información de SyE comisionada por las partes interesadas clave (por ejemplo, el Ministerio de Finanzas) es utilizada por otros (ministerios sectoriales). Identificación de barreras para el uso (si las hay). Evidencia referente al grado de utilización de las distintas partes interesadas (por ejemplo, una revisión de diagnóstico o una encuesta). Ejemplos de evaluaciones importantes que hayan sido muy influyentes con el gobierno.
7. Tipos de herramientas de mayor utilización en el sistema de SyE (indicadores del desempeño regular, revisiones o evaluaciones rápidas, auditorías de desempeño, evaluaciones de impacto profundas y rigurosas). Escala y costo de cada uno de estos tipos de SyE. La manera como se fijan las prioridades de la evaluación (¿se centran en "programas problema", programas piloto, programas de alto gasto o alta visibilidad? ¿O se basan en una agenda de investigación sistemática para responder preguntas sobre la eficacia del programa?).
8. Responsabilidad de recopilar información sobre el desempeño y realizar evaluaciones (los ministerios, la academia o las empresas consultoras). Problemas con la calidad o la confiabilidad de los datos o con la calidad de las evaluaciones que se han realizado. Fortalezas y

(Continúa en la página siguiente)

(Continuación recuadro 15.1)

debilidades de la oferta local de SyE. Restricciones básicas de capacidad y prioridades en la generación de capacidad del gobierno.

9. Grado de apoyo de los donantes para el SyE en los últimos años. Proyectos de donantes que apoyan el SyE a nivel de gobierno general, sectorial o de agencia (provisión de asistencia técnica, otra generación de capacidad y financiación para evaluaciones importantes, como las de impacto riguroso).
10. Conclusiones sobre las fortalezas y debilidades en general del sistema de SyE, incluyendo su sostenibilidad frente a un cambio de gobierno. Dependencia del sistema de la financiación de los donantes o de otro apoyo. Planes corrientes para el futuro fortalecimiento del sistema de SyE.

Fuente: Banco Mundial, 2006.

agencias de ayuda para el desarrollo en un enfoque de gobierno total para el desarrollo aumenta la compensación por tener medidas nacionales de desempeño disponibles.

El desarrollo de la capacidad de evaluación requiere convencer a los autores de políticas de que la evaluación es útil; no puede imponérsele a un gobierno. Requiere una infraestructura de información que permita una recopilación rutinaria de datos y la divulgación y el uso de los resultados. Generar sistemas de evaluación útiles toma usualmente un largo tiempo, ya que requiere cambios políticos e institucionales (Schaumberg-Müller, 1996). Por ejemplo, a los Estados Unidos le tomó 10 años implementar totalmente el Government Performance Results Act (GPRA) y todos los sistemas de SyE dentro de ella (Kusek y Rist, 2004). Las instituciones necesitan creer que las evaluaciones no son sistemas de seguimiento, sino herramientas para mejorar el desempeño y la toma de decisiones.

¿Qué puede hacerse para motivar el desarrollo de la capacidad de evaluación? ¿Qué incentivos y restricciones existen? Toulemonde (1999) discute la importancia de utilizar tanto incentivos (zanahorias) como evaluación de la demanda (garrotes). El primer incentivo que identifica es presupuestario. Las evaluaciones cuestan dinero que debe estar disponible para desarrollar capacidad de evaluación. El segundo incentivo es una buena trayectoria de carrera para quienes eligen ser evaluadores.

Toulemonde identifica formas de crear demanda de evaluaciones mediante restricciones. La primera es la evaluación obligatoria. Cuando se requiere a las organizaciones realizar evaluaciones, debe otorgárseles el derecho de hacer preguntas. Deben tener acceso completo a datos de campo y exigírseles hacer uso de los resultados de las evaluaciones.

Toulemonde admite que "las leyes y las reglas no tienen significado en sí mismas" (1999, p. 159). Deben hacerse cumplir y ejercerse el poder. Él continúa discutiendo el valor de combinar los incentivos y las restricciones.

Profesionalismo de los evaluadores para el desarrollo

Cualquiera puede decidir asumir la identidad de evaluador para el desarrollo. Todavía no existen competencias estándar que deban poseerse mediante capacitación o práctica; tampoco, proceso de credencial o certificación que se necesite completar para llamarse a sí mismo evaluador para el desarrollo. Muchas personas aprenden en el trabajo sin capacitación formal. Administrar o dirigir unidades de evaluación tampoco requiere credenciales especiales: el conocimiento para evaluar y la experiencia no constituyen requisitos previos.

El método demostrado con el tiempo de aprender en el trabajo no debe descontarse; los administradores sin capacitación formal pueden progresar en el trabajo. La falta de una credencial, sin embargo, dificulta identificar la pericia a quienes adquieren servicios de evaluación para el desarrollo.

Comoquiera que la demanda por evaluación para el desarrollo continúa creciendo, pensamos que las crecientes exigencias de profesionalismo son probables como respuesta a evaluaciones concebidas o realizadas deficientemente. La Asociación Internacional de Desarrollo de Evaluación (IDEAS) está representando un papel cada vez más importante en la profesionalización mundial de la evaluación para el desarrollo.

Fijar competencias fundamentales

Un tema importante dentro y fuera de la evaluación para el desarrollo es el de si deben fijarse competencias centrales para un evaluador, como las presentadas en el recuadro 15.2. La AEA, después de una investigación inicial, ha elegido no respaldar dichas competencias. En contraste, la Canadian Evaluation Society (CES) está persiguiendo las "designaciones profesionales", una forma de otorgar credenciales y, simultáneamente, está buscando incrementar capacitación con grado en evaluación. La New Zealand Evaluation Association está actualmente considerando el desarrollo de competencias.

La Treasury Board de Canadá está exigiendo certificación para los jefes de las unidades de evaluación que hayan sido ordenadas por agencias canadienses. En un informe preparado por su Secretaría se presentan los resultados de un estudio para "investigar enfoques opcionales para aumentar el profesionalismo de la evaluación en el gobierno federal" (Treasury of Board of Canada Secretariat, 2005). La International Board of Standards for Training, Performance, and Instruction ha desarrollado competencias para personal interno y consultores externos que realicen evaluaciones en entornos organizativos (el vínculo de este sitio web se encuentra al final de este capítulo).

En el CES's Professional Designations Project se piensa que la credencial debe estar basada en una combinación de educación y experiencia, y se considera que la certificación requiere un examen, por lo que se está escogiendo un enfoque de credencial no basada en examen. Muchas de las tareas requeridas para establecer la credencial están en camino, incluidos el establecimiento de requerimientos para la credencial, los

mecanismos de *grand parenting* para quienes ya son practicantes profesionales, una junta de otorgamiento de credenciales y un mecanismo de disputa. Se espera indicar un proyecto de demostración en 2009 (AEA, 2008). El vínculo de la CES, al final de este capítulo, presenta mayor información sobre este esfuerzo.

Recuadro 15.2 *Competencias esenciales para evaluadores de programas*

1.0 Práctica profesional:
1.1 Aplica estándares de evaluación profesional.
1.2 Actúa con ética y se esfuerza por la integridad y la honestidad en la realización de las evaluaciones.
1.3 Transmite métodos y destrezas personales de evaluación a los clientes potenciales.
1.4 Respeta a los clientes, encuestados, participantes en los programas y a otras partes interesadas.
1.5 Tiene en cuenta el bienestar general y público en la práctica de la evaluación.
1.6 Contribuye a la base de conocimientos sobre evaluación.

2.0 Investigación sistemática:
2.1 Conoce la base de conocimientos sobre evaluación (términos, conceptos, teorías, supuestos).
2.2 Conoce los métodos cuantitativos.
2.3 Conoce los métodos cualitativos.
2.4 Conoce los métodos mixtos.
2.5 Realiza revisiones de la literatura.
2.6 Especifica la teoría del programa.
2.7 Enmarca las preguntas de evaluación.
2.8 Desarrolla diseños de evaluación.
2.9 Identifica las fuentes de datos.
2.10 Recopila datos.
2.11 Evalúa la validez de los datos.
2.12 Evalúa la confiabilidad de los datos.
2.13 Analiza los datos.
2.14 Interpreta los datos.
2.15 Hace juicios.
2.16 Hace recomendaciones.
2.17 Aporta fundamentos racionales para las decisiones a través de la evaluación.
2.18 Informa los procedimientos y resultados de la evaluación.
2.19 Anota las fortalezas y limitaciones de la evaluación.
2.20 Realiza metaevaluaciones.

3.0 Análisis situacional:
3.1 Describe el programa
3.2 Determina la evaluabilidad del programa.

(Continúa en la página siguiente)

(Continuación recuadro 15.2)

3.3 Identifica los intereses de las partes interesadas pertinentes.
3.4 Atiende las necesidades de información de los usuarios previstos.
3.5 Trabaja sobre los conflictos.
3.6 Examina el contexto organizativo de la evaluación.
3.7 Analiza las consideraciones políticas pertinentes a la evaluación.
3.8 Presta atención a las cuestiones del uso de la evaluación.
3.9 Presta atención a las cuestiones de cambios en la organización.
3.10 Respeta la singularidad del sitio de la evaluación y del cliente.
3.11 Permanece abierto a los aportes de otras personas.
3.12 Modifica el estudio cuando sea necesario.

4.0 Administración del proyecto:
4.1 Responde a solicitudes de propuestas.
4.2 Negocia con el cliente antes de iniciar la evaluación.
4.3 Elabora los acuerdos formales.
4.4 Mantiene comunicación con el cliente durante todo el proceso de evaluación.
4.5 Presupuesta una evaluación.
4.6 Justifica el costo según las necesidades de la información.
4.7 Identifica los recursos necesarios para la evaluación, tales como información, pericia, personal e instrumentos.
4.8 Utiliza tecnologías apropiadas.
4.9 Supervisa a otras personas incluidas en la realización de la evaluación.
4.10 Capacita a otras personas incluidas en la realización de la evaluación.
4.11 Realiza la evaluación en forma no destructiva.
4.12 Presenta el trabajo oportunamente.

5.0 Práctica reflexiva:
5.1 Tiene conciencia de sí mismo como evaluador (conocimiento, destrezas, disposición).
5.2 Reflexiona sobre la práctica de evaluación personal (competencias y campos de crecimiento).
5.3 Busca el desarrollo profesional en evaluación.
5.4 Busca el desarrollo profesional en áreas de contenido pertinente.
5.5 Construye relaciones profesionales para enriquecer la práctica de la evaluación.

6.0 Competencia interpersonal:
6.1 Utiliza destrezas de comunicación escrita.
6.2 Utiliza destrezas de comunicación verbal y de escucha.
6.3 Utiliza destrezas de negociación.
6.4 Utiliza destrezas de resolución de conflictos.
6.5 Facilita la interacción interpersonal constructiva (trabajo en equipo, facilitación de grupos, procesamiento).
6.6 Demuestra competencia intercultural.

Fuente: Stevahn y otros, 2005.

¿Qué hay sobre los estándares para quienes se ocupan en la evaluación para el desarrollo? Ideas está empezando a explorar este asunto para los evaluadores en el desarrollo. Los esfuerzos extensos y el trabajo de antecedentes de la Canadian Evaluation Society constituirán un punto de partida. El trabajo de la Evaluation Capacity Development Task Force del UNEG –que desarrollaron competencias centrales para el personal de evaluación en diferentes niveles y fueron adoptados en 2008– puede ser también útil (el vínculo del sitio web de UNEG se encuentra al final de este capítulo).

Algunos creen que en el campo de la evaluación nunca se estará de acuerdo sobre un conjunto de competencias para los evaluadores y que, en cambio, cada organización u otra entidad desarrollará competencias para los evaluadores, específicas de su negocio, cultura y tradiciones. Pronosticamos que se desarrollará un conjunto de competencias centrales para la evaluación para el desarrollo. Probablemente las organizaciones tendrán competencias adicionales para los evaluadores, específicas de su actividad, y las adecuarán, pero esperarán a que posean las competencias centrales.

Otro argumento es el de que se tendría que ser un superhumano para poseer todas las competencias necesarias para todas las evaluaciones en todas las situaciones. Ciertamente, habrá requisitos adicionales para quienes buscan especialidades de evaluación, como pericia en evaluación social y ambiental u otras parecidas. Sin embargo, así como un cirujano del cerebro se capacita en medicina general, pero los médicos no se capacitan en cirugía cerebral, el núcleo es central.

Certificación para evaluadores para el desarrollo

¿Debería haber un proceso mediante el cual los evaluadores para el desarrollo puedan adquirir certificación de destrezas reconocida mundialmente? Tal certificación ha sido ciertamente útil para los auditores, entre quienes la designación como contador público certificado es reconocida y respetada internacionalmente. Una encuesta reciente de Ideas a sus afiliados indicó un fuerte apoyo para una mayor profesionalización a través de competencias y, a la larga, la certificación (Morra Imas, 2009).

La certificación para los evaluadores proporcionaría un mecanismo para proteger a los consumidores del trabajo por debajo de los estándares y los profesionales que falazmente embellecen sus credenciales. También es un mecanismo para distinguir la evaluación y a quienes la realizan de las revisiones de auditorías y los auditores. Las preocupaciones se centran en temores de que crear un mecanismo podría crear una barrera a la entrada, especialmente en un entorno donde la demanda supera la oferta, y en un campo donde el aprendizaje en el trabajo ha constituido un método de entrada tradicional.

Donde la calidad de la evaluación y la competencia del evaluador constituyen una preocupación, especialmente por la alta demanda externa, es probable una presión fuerte por competencias básicas. Nuestra opinión es que si se establecen las competencias centrales para los evaluadores para el desarrollo, la discusión sobre la certificación no estará muy distante.

Al final, los mayores desafíos en la evaluación para el desarrollo pueden ser aprender a fijar expectativas realistas para las intervenciones, generar demanda pública por evaluación en países sin tradiciones democráticas, reconciliar la evaluación basada en resultados con la evaluación participativa y promover el pensamiento evaluativo dentro de las organizaciones. Esperamos que estos desafíos se superen.

Actividad del capítulo 15

Ejercicio de aplicación 15.1: generar capacidad de evaluación

El gobierno le ha solicitado a usted (su grupo) desarrollar un plan estratégico para incrementar la capacidad de evaluación en su país. Para orientar sus ideas de preparar un plan estratégico de una a dos páginas, utilice las preguntas siguientes:

1. ¿Cuáles son los dos o tres temas de desarrollo más difíciles que deben atacarse en los próximos años?
2. De lo que usted conoce. ¿qué capacidad de evaluación existe ya (disponibilidad de evaluadores, destrezas, recursos, infraestructura)?
3. Dadas las necesidades actuales y futuras y los problemas y su valoración de la capacidad actual de evaluación, enumere las seis mejoras más importantes que serían más provechosas para la capacidad de evaluación en su país.
4. ¿Qué está impulsando la necesidad de sistemas de SyE en su país?
5. ¿Dónde recae en su gobierno la responsabilidad por la ejecución eficaz (y eficiente) de los programas?
6. ¿Existe una estrategia codificada (estatutos o mandato) u organización en el gobierno para hacer seguimiento a los objetivos de desarrollo?
7. ¿Dónde reside la capacidad con las destrezas requeridas para diseñar y utilizar sistemas de SyE en su país? ¿Cómo ha contribuido esta capacidad (o la falta de ella) al empleo de SyE en su país?
8. Fije prioridades a las necesidades de su país, identificándolas como críticas, muy importantes o importantes.

Referencias y lecturas adicionales

AEA (American Evaluation Association). 2008. "Professional Designation for Evaluators in Canada: Where Are We Now?" Sesión de la conferencia anual de la American Evaluation Association, Denver, CO, noviembre 5-8.

Banco Mundial. 1999. *Evaluation Capacity Development: A Diagnostic Guide and Action Framework.* ECD, Documento de trabajo 6, Operations Evaluation Department, Washington, DC.

———. 2006. *Diagnostic Guides.* Washington, DC. http://www.worldbank.org/ieg/ecd/diagnostic_guides.html.

Beywl, W. y K. Harich. 2008. "University-Based Continuing Education in Evaluation: The Baseline in Europe". *Evaluation* 13 (1): 121-34.

Boyle, Richard. 1999. "Professionalizing the Evaluation Function: Human Resource Development and the Building of Evaluation Capacity". En *Building Effective Evaluation Capacity: Lessons from Practice,* eds. Richard Boyle y Donald Lemaire, 135-51. Nueva Brunswick, NJ: Transaction Publishers.

Boyle, Richard y Donald Lemaire, eds. 1999. *Building Effective Evaluation Capacity: Lessons from Practice.* Nueva Brunswick, NJ: Transaction Publishers.

Dewey, J. D., B. E. Montrosse, D. C. Schroter, C. D. Sullins y J. R. Mattox II. 2008. "Evaluator Competencies: What's Taught versus What's Sought". *American Journal of Evaluation* 29 (3): 268-87.

Ghere, G., J. A. King, L. Stevahn y J. Minnema. 2006. "A Professional Development Unit for Reflecting on Program Evaluator Competencies". *American Journal of Evaluation,* 27 (1): 108-23.

Kusek, Jody Zall y Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System.* Banco Mundial, Washington, DC. http://www.oecd.org/dataoecd/23/27/35281194.pdf.

Mackay, Keith. 2007. *How to Build M&E Systems to Support Better Government.* Banco Mundial, Independent Evaluation Group, Washington, DC. http://www.worldbank.org/ieg/ecd/docs/How_to_build_ME_gov.pdf.

McDonald, K. E. y S. E. Myrick. 2008. "Principles, Promises y a Personal Plea: What Is an Evaluator to Do?" *American Journal of Evaluation* 29 (3): 343-51.

Morra Imas, Linda. 2009. "The Competencies and Credentials Debate in Evaluation: What Is It all About?" Presentación para la International Development Evaluation Association's Global Assembly, Johannesburgo, marzo.

Porteous, Nancy L., Barbara J. Sheldrick y Paula J. Stewart. 1999. "Enhancing Managers' Evaluation Capacity: A Case Study for Ontario Public Heath". *Canadian Journal of Program Evaluation* (edición especial): 137-54.

Rist, Ray C. y Nicoletta Stame, eds. 2006. From Studies to Streams: *Managing Evaluative Systems.* Piscataway, NJ: Transaction Publishers.

Schaumberg-Müller, Henrik. 1996. *Evaluation Capacity Building: Donor Support and Experiences.* Copenhague: Organisation for Economic Co-operation and Development. http://www.oecd.org/dataoecd/20/52/16546669.pdf.

Sonnichsen, Richard C. 1999. "Building Evaluation Capacity within Organizations". En *Building Effective Evaluation Capacity: Lessons from Practice,* eds. Richard Boyle y Donald Lemaire, 53-73. New Brunswick, NJ: Transaction Publishers.

Stevahn, Laurie, Jean A. King, Gail Ghere y Jane Minnema. 2005. "Establishing Essential Competencies for Program Evaluators". *American Journal of Evaluation* 26 (1): 43-59.

Toulemonde, Jacques. 1999. "Incentives, Constraints y Culture-Building as Instruments for Development of Evaluation Demand". En *Building Effective Evaluation Capacity: Lessons from Practice,* eds. Richard Boyle y Donald Lemaire, 153-68. New Brunswick, NJ: Transaction Publishers.

Treasury of Board of Canada Secretariat. 2005. *Improving the Professionalism of Evaluation.* Informe final. http://www.tbs-sct.gc.ca/eval/dev/Professionalism/profession_e.asp.

Sitios web

Campbell, Patricia B. y Beatriz Chu Clewell. 2008. *Building Evaluation Capacity: Designing a Cross-Project Evaluation. Guide I.* Urban Institute, Washington, DC. http://www.uquebec.ca/observgo/fichiers/43799_GRA-1a.pdf.

———. 2008. *Building Evaluation Capacity: Collecting and Using Data in Cross-Project Evaluations. Guide II.* Urban Institute, Washington, DC. http://www.uquebec.ca/observgo/fichiers/42773_GRA-1b.pdf.

Canadian Evaluation Society. www.evaluationcanada.ca.

———. *Professional Designation Project.* http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=EN.

UNEG (Grupo de Evaluación de las Naciones Unidas).2004a. *Evaluation Capacity Development: A Strategy Paper.* http://cfapp1-docs-public.undp.org/eo/evaldocs1/uneg_2006/eo_doc_350011048.doc.

———. 2004b. *Evaluation Capacity Development Task Force: Highlights.* http://cfapp1-docs-public.undp.org/eo/evaldocs1/uneg_2006/eo_doc_562044935.doc.

Weiss, Carol. *Evaluating Capacity Development: Experiences from Research and Development Organizations around the World.* http://www.agricta.org/pubs/isnar2/ECDbood(H-ch7).pdf.

APÉNDICES

001El camino.indd 499 08/11/2010 03:28:02 p.m.

001El camino.indd 500 08/11/2010 03:28:02 p.m.

Red de la OCDE CAD sobre la evaluación del desarrollo. Evaluación de la cooperación para el desarrollo: sumario de normas y estándares básicos

Contenido

Introducción	502
Definición de evaluación	503
Desarrollo de normas y estándares para evaluación	503
Parte I: Principios del CAD para la evaluación de asistencia para el desarrollo	504
Revisión de los Principios del CAD para evaluación de asistencia para el desarrollo	507
Parte II: Criterios de evaluación	508
Parte III: Sistemas de evaluación y uso: herramienta de trabajo para revisiones y evaluaciones de pares	509
Parte IV: Estándares de calidad de evaluación del CAD	512
Parte V: Documentos de orientación	516

001El camino.indd 501 08/11/2010 03:28:02 p.m.

La Red del CAD sobre Evaluación para el Desarrollo es un órgano subsidiario del Development Assistance Committee (CAD), cuyo propósito es incrementar la eficacia de los programas internacionales para el desarrollo mediante el apoyo a la evaluación robusta, informada e independiente. La Red es un órgano único que agrupa a 30 donantes bilaterales y agencias de desarrollo multilateral: Australia, Austria, Bélgica, Canadá, Dinamarca, Comisión Europea, Finlandia, Francia, Alemania, Grecia, Irlanda, Italia, Japón, Luxemburgo, Países Bajos, Nueva Zelanda, Noruega, Portugal, España, Suecia, Suiza, Reino Unido, Estados Unidos, Banco Mundial, Asian Development Bank, African Development Bank, Banco Interamericano de Desarrollo, Banco Europeo de Reconstrucción y Desarrollo, PNUD y el FMI.

Introducción

La Red del CAD sobre Evaluación para el Desarrollo es un foro internacional único que agrupa a directores de evaluación y especialistas de agencias de cooperación para el desarrollo, en los países integrantes de la OCDE, y a instituciones de desarrollo multilateral. Su meta es incrementar la eficacia de los programas internacionales para el desarrollo, apoyando la evaluación robusta, informada e independiente. La Red es un órgano subsidiario del Development Assistance Committee (CAD).

Un componente clave de la misión de la Red es desarrollar normas y estándares acordados internacionalmente en la evaluación para el desarrollo, las cuales constituyen información básica para la política y la práctica de la evaluación, y contribuyen a la armonización de los enfoques de acuerdo con los compromisos de la Declaración de París sobre Eficacia de la Ayuda. El cuerpo de normas y estándares está basado en la experiencia y evoluciona con el tiempo para adaptarse al entorno cambiante de la ayuda para el desarrollo. Las normas y estándares sirven como punto de referencia internacional, orientando los esfuerzos para mejorar los resultados del desarrollo a través de la evaluación de alta calidad.

Las normas y los estándares que aquí se resumen deben aplicarse juiciosamente y adaptarse de manera cuidadosa al propósito, al objeto y al contexto de cada evaluación. Dado que este documento sumario no constituye un manual exhaustivo de evaluación, se alienta a los lectores a consultar los textos completos que se encuentran disponibles en sitio web de la Red del CAD sobre Evaluación para el Desarrollo:

www.oecd.org/dac/evaluationnetwork.

Esta y otras definiciones básicas están incluidas en el *Glossary of Key Terms in Evaluation and Results Based Management*. El glosario es una útil herramienta de desarrollo de capacidades que contribuye a la comprensión de los conceptos fundamentales de la evaluación. El glosario del CAD está disponible en los idiomas árabe, chino, holandés, inglés, francés, italiano, japonés, kiswahili, portugués, ruso, español, sueco y turco.

Definición de evaluación

La evaluación es la valoración sistemática y objetiva de un proyecto, un programa o una política en curso o finalizado, su diseño, implementación y resultados.

Su propósito es establecer la pertinencia y el cumplimiento de los objetivos, la eficiencia, la eficacia, el impacto y la sostenibilidad para el desarrollo. Una evaluación debe suministrar información verosímil y útil, habilitando la incorporación de las enseñanzas aprendidas en el proceso de toma de decisiones de receptores y donantes.

La evaluación se refiere también al proceso de determinar el valor o la significación de una actividad, una política o un programa.

Desarrollo de normas y estándares para evaluación

Un conjunto de *principios para la evaluación de la asistencia para el desarrollo* (que se resumen en la parte I) fue adoptado por la OCDE CAD en 1991, y estos principios constituyen el centro del enfoque de evaluación de la Evaluation Network. Los principios se centran en la administración y en acuerdos institucionales del sistema de evaluación dentro de las agencias para el desarrollo.

Durante una *revisión de los principios de evaluación* en 1998, la mayoría de los integrantes del CAD informaron haber realizado progreso, implementando los principios centrales, y los encontraron útiles y pertinentes. Estos principios fundamentales de evaluación no solamente continúan como una norma de referencia clave para la evaluación para el desarrollo, sino que sirven también como base para las CAD Peer Reviews, el único mecanismo acordado internacionalmente para evaluar el desempeño de los programas de cooperación para el desarrollo de la OCDE CAD. Sin embargo, la revisión destacó también áreas que requieren ajuste o una orientación más específica, preparando el camino para desarrollos adicionales.

Los *criterios del CAD para la evaluación de asistencia para el desarrollo* (detallados en la Parte II) se basan en los principios de evaluación, y sirven como guía general de medidas de aplicación y son útiles para evaluar el trabajo sobre el desarrollo.

Un análisis completo de las políticas de evaluación de los integrantes emprendido en 2006, basado en una revisión de revisiones pares efectuada en un período de ocho años, llevó al desarrollo de los *Evaluation Systems and Use: A Working Tool for Peer Reviews and Assessments* (Parte III). Este documento proporciona los elementos básicos de una función sólida de evaluación en las agencias para el desarrollo, y se utiliza para avanzar la implementación de los principios.

El paso más reciente en el desarrollo de un marco normativo es la definición de *estándares de calidad de evaluación* (presentados en borrador en la Parte IV). Estos estándares ofrecen orientación sobre los procesos y productos de las evaluaciones, y quedarán terminados en 2009-2010, después de una fase de prueba de tres años.

Además de estas normas y estándares generales, los integrantes del OECD CAD reconocen la necesidad de guías específicas en ciertos campos de la evaluación para el desarrollo. Con base en las experiencias de evaluación y los textos antes descritos, se ha desarrollado la guía en varios campos. Los *documentos guía* más significativos se presentan en la Parte V.

Parte I: Principios del CAD para la evaluación de asistencia para el desarrollo

Los principios de evaluación fueron adoptados en la Reunión de Alto Nivel del OCDE CAD en 1991 y publicados en 1992 como parte de los *CAD Principles for Effective Aid*. A continuación se presenta una visión general de los elementos básicos del documento original.

I. Mensajes centrales

Los principios proporcionan una orientación general sobre el papel de la evaluación de la ayuda en el proceso de su administración, con los siguientes mensajes centrales:

- Las agencias de ayuda deben tener una política de evaluación con lineamientos y métodos claros y con una definición clara de su papel, responsabilidades y su lugar en la estructura de la ayuda institucional.
- El proceso de evaluación debe ser imparcial e independiente del proceso de la formulación de políticas, y de la ejecución y manejo de la ayuda para el desarrollo.
- El proceso de evaluación debe ser lo más abierto posible y divulgar los resultados ampliamente.
- Para que las evaluaciones resulten útiles, deben utilizarse. Es esencial la retroalimentación a los autores de políticas y al personal operativo.
- La asociación con los receptores y la cooperación de los donantes en la evaluación de la ayuda son esenciales, y constituyen un aspecto importante de la formación institucional de los receptores y de la coordinación de la ayuda, y reducen la carga administrativa de los receptores.
- La evaluación de la ayuda y sus requerimientos debe ser parte integral de la planeación de la ayuda desde el principio. La identificación clara de los objetivos que deben lograrse con una actividad de ayuda es un prerrequisito esencial para la evaluación objetiva. (Par. 4).

II. Propósito de la evaluación

Los principales propósitos de la evaluación son:

- Mejorar las políticas, los programas y proyectos futuros de la ayuda, mediante la retroalimentación de las enseñanzas aprendidas.
- Proporcionar una base para la rendición de cuentas, incluyendo el suministro de información al público.

A través de la evaluación tanto de fracasos como de éxitos se genera información valiosa que, de retroalimentarse de manera adecuada, puede mejorar los programas y proyectos futuros de ayuda. (Par. 6).

III. Imparcialidad e independencia

El proceso de evaluación debe ser imparcial e independiente de los procesos de formulación de políticas y de la ejecución y administración de la ayuda para el desarrollo. (Par. 11).

La imparcialidad contribuye a la credibilidad de la evaluación y a evitar sesgos en los hallazgos, análisis y conclusiones. La independencia proporciona legitimidad a la evaluación y reduce el potencial de conflicto de intereses que podría surgir si los autores de políticas y administradores fueran solamente responsables de evaluar sus propias actividades. (Par. 12).

La imparcialidad y la independencia se obtienen mejor separando la función de evaluación de la administración de línea responsable de la planeación, y la administración de la ayuda para el desarrollo. Esto puede lograrse con una unidad central responsable de la evaluación que reporte directamente al ministro o jefe de agencia responsable de la asistencia para el desarrollo, o a una junta de directores o gobernadores de la institución. Según el grado en que algunas funciones de evaluación se adjunten a la administración de línea, deberían reportar a una unidad central o a un nivel suficientemente alto de la estructura administrativa, o a un comité de administración responsable de las decisiones del programa. En este caso, deben hacerse todos los esfuerzos para evitar comprometer el proceso de evaluación y sus resultados. Con cualquier enfoque que se seleccione, los acuerdos y procedimientos organizativos deben facilitar el vínculo de los hallazgos de la evaluación con la programación y la formulación de políticas. (Par. 16).

IV. Credibilidad

La credibilidad de la evaluación depende de la pericia y la independencia de los evaluadores y el grado de transparencia del proceso de evaluación. La credibilidad exige que la evaluación informe sobre los éxitos y los fracasos. Los países receptores deberían, como regla general, participar totalmente en la evaluación a fin de promover la credibilidad y el compromiso. (Par. 18).

La transparencia del proceso de evaluación es crucial para la credibilidad y la legitimidad… (Par. 20).

V. Utilidad

Para causar impacto en la toma de decisiones, los hallazgos de la evaluación deben percibirse como aplicables y útiles y presentarse en forma clara y concisa. Deben reflejar los distintos intereses y necesidades de las muchas partes interesadas involucradas en la cooperación para el desarrollo. Una fácil accesibilidad es también crucial para la utilidad… (Par. 21).

Las evaluaciones deben ser oportunas en el sentido de que deben estar disponibles en un momento apropiado para el proceso de toma de decisiones. Esto sugiere que la evaluación tiene un papel importante en distintas etapas durante la ejecución de un proyecto o un programa, y no debe realizarse solamente como un ejercicio *ex post*. El personal operativo es responsable del seguimiento de las actividades en progreso. Las provisiones para la evaluación por parte de personal independiente en el plan de operación constituyen un complemento importante para regular el seguimiento. (Par. 22).

VI. Participación de donantes y receptores

… en cuanto sea posible, tanto los donantes como los receptores deben participar en el proceso de evaluación. Como los hallazgos de la evaluación son pertinentes a ambas partes, los términos de referencia de la evaluación deben tratar asuntos de interés para cada parte, y la evaluación debe

001El camino.indd 505 08/11/2010 03:28:02 p.m.

reflejar sus opiniones sobre la eficacia y el impacto de las actividades pertinentes. El principio de imparcialidad y de independencia durante la evaluación debe aplicarse por igual a receptores y donantes. La participación y la imparcialidad mejoran la calidad de la evaluación, lo que a su vez tiene implicaciones significativas para la sostenibilidad a largo plazo, ya que los receptores son los únicos responsables después de la partida del donante. (Par. 23).

Cuando sea apropiado, las opiniones y la pericia de los grupos afectados deben formar parte integral de la evaluación. (Par. 24).

La participación de todas las partes ofrece la oportunidad de aprender haciendo, y fortalece las destrezas y capacidades en los países receptores, lo que constituye un objetivo importante que debe también promoverse mediante la capacitación y demás apoyo al desarrollo institucional y administrativo. (Par. 25).

VII. Cooperación de los donantes

La colaboración entre los donantes es esencial para el aprendizaje mutuo y para evitar la duplicación de esfuerzos. La colaboración de los donantes debe estimularse, a fin de desarrollar métodos de evaluación, compartir informes e información y mejorar el acceso a los hallazgos de la evaluación. Las evaluaciones de donantes conjuntos deben promoverse, para mejorar el conocimiento de los procedimientos y enfoques de los demás y reducir la carga administrativa sobre el receptor. Con el propósito de facilitar la planeación de evaluaciones conjuntas, los donantes deben intercambiar planes de evaluación sistemáticamente y con una buena anticipación a la implementación. (Par. 26).

VIII. Programación de la evaluación

La agencia debe desarrollar un plan general para la evaluación de las actividades de asistencia para el desarrollo. Al elaborarlo, deben organizarse por categorías apropiadas las distintas actividades que se van a evaluar. Deben fijarse luego las prioridades para la evaluación de las categorías y prepararse un cronograma. (Par. 27).

Las agencias de ayuda que no lo hayan hecho, deben elaborar lineamientos y/o estándares para el proceso de evaluación, los cuales deben proveer orientación y definir los requisitos mínimos para la realización de evaluaciones y la presentación de información. (Par. 31).

IX. Diseño e implementación de evaluaciones

Cada evaluación debe planearse y emplear términos de referencia para:

- Definir el propósito y el alcance de la evaluación, incluyendo una identificación de los receptores de los hallazgos.
- Describir los métodos que se utilizarán en la evaluación.
- Identificar los estándares frente a los cuales se evaluará el desempeño del proyecto o el programa.
- Determinar los recursos y el tiempo requerido para completar la evaluación. (Par. 32).

001El camino.indd 506 08/11/2010 03:28:02 p.m.

Es esencial definir las preguntas que se deberán responder en la evaluación, que a menudo se conocen como los "temas" de la evaluación. Los temas proveen un marco manejable para el proceso de evaluación y la base para un conjunto claro de conclusiones y recomendaciones… (Par. 35).

X. Divulgación de informes y retroalimentación

La información de la evaluación debe ser clara, tan libre como sea posible de lenguaje técnico, e incluir los elementos siguientes: un resumen ejecutivo; un perfil de la actividad evaluada; una descripción de los métodos de evaluación utilizados; los hallazgos principales; las enseñanzas aprendidas; las conclusiones y recomendaciones (que deben estar separadas del informe mismo). (Par. 39).

La retroalimentación es parte esencial del proceso de evaluación, ya que provee el vínculo entre las actividades pasadas y las futuras. Para asegurar que los resultados de las evaluaciones se utilicen en las políticas y desarrollo de programas futuros, es necesario establecer mecanismos de retroalimentación que incluyan a todas las partes interesadas. Estos mecanismos incluyen medidas, como comités de evaluación, seminarios y talleres, sistemas automatizados, procedimientos de información y seguimiento. Los medios informales, por ejemplo redes y comunicaciones internas, permitirían también la divulgación de ideas y de información. Para ser eficaz, el proceso de retroalimentación debe contar con recursos de personal y presupuestarios, al igual que apoyo de la administración principal y otros actores. (Par. 42).

Revisión de los Principios del CAD para evaluación de asistencia para el desarrollo

En 1998, los integrantes del Working Party on Aid Evaluation (ahora la CAD Network on Development Evaluation) finalizaron una revisión de su experiencia con la aplicación de los *CAD Principles for Evaluation of Development Assistance*. El objetivo fue examinar la implementación y el uso de los principios, con el fin de evaluar su impacto, utilidad y pertinencia y hacer recomendaciones. El resumen suministrado más adelante demuestra los esfuerzos actuales que se hacen para implementar los principios, y señalan el camino hacia algo del trabajo posterior presentado en las partes II a V de este documento. El texto completo incluye hallazgos detallados y recomendaciones adicionales de evaluadores y usuarios.

La revisión demostró que la evaluación en la cooperación para el desarrollo está evolucionando y cambiando su enfoque. La mayor parte de los miembros de la Network han reorganizado sus oficinas centrales de evaluación, para tener un nuevo papel con un enfoque más fuerte en la eficacia de la ayuda. Además, las oficinas centrales de evaluación se han alejado de la tradicional evaluación de proyectos a favor de evaluaciones de programas, sectores, temáticas y de ayuda a países. En los países de la OCDE ha crecido el interés interno en los resultados de la ayuda para el desarrollo. Igualmente, se ha reportado un mayor interés en las evaluaciones descentralizadas, y existe evidencia que sugiere que la evaluación en los países en desarrollo está empezando a echar raíces.

La mayoría de los miembros informó que ha logrado un buen grado de cumplimiento con los *CAD Principles for Evaluation of Development Assistance* esenciales, y también sostienen ha-

001El camino.indd 507 08/11/2010 03:28:02 p.m.

berlos encontrado útiles y pertinentes para orientar su trabajo y, en algunos casos, reorganizar sus oficinas de evaluación. Con base en estos resultados se concluyó que los principios todavía son válidos y sensatos.

No obstante, se reconoció que los principios deben complementarse y reforzarse con orientación (por ejemplo, buenas o mejores prácticas) y áreas clave. Incluyen formas de: manejar eficazmente las compensaciones entre independencia y participación requerida para obtener asociación; mejorar la retroalimentación y las prácticas de comunicación; promover una cultura de evaluación; implementar evaluaciones de programas de país y conjuntas; promover asociaciones y evaluar la ayuda humanitaria.

Parte II: Criterios de evaluación

Al evaluar programas y proyectos de cooperación para el desarrollo, es útil considerar los siguientes criterios, presentados en los *CAD Principles for Evaluation of Development Assistance*.

Pertinencia

El grado al cual la actividad de ayuda se ajusta a las prioridades y políticas del grupo objetivo, receptor y donante.

Al evaluar la pertinencia de un programa o un proyecto, es útil considerar las preguntas siguientes:

- ¿Hasta qué punto son todavía válidos los objetivos del programa?
- ¿Son consistentes las actividades y productos del programa con la meta general y el logro de sus objetivos?
- ¿Son consistentes las actividades y productos del programa con los impactos y efectos previstos?

Eficacia

Una medida del grado al cual una actividad de ayuda logra sus objetivos.

Al evaluar la eficacia de un programa o un proyecto, es útil considerar las preguntas siguientes:

- ¿Hasta qué punto se lograron o es probable que se logren los objetivos?
- ¿Cuáles fueron los principales factores que influyeron en el logro o no logro de los objetivos?

Eficiencia

La eficiencia mide los productos –cualitativos y cuantitativos– en relación con los insumos. Es un término económico utilizado para evaluar el grado al cual la ayuda para el desarrollo utiliza los recursos menos costosos posibles para lograr los resultados deseados. Esto generalmente requiere comparar enfoques alternativos para obtener los mismos productos, y verificar si se ha adoptado el proceso más eficiente.

Al evaluar la eficiencia de un programa o un proyecto es útil considerar las preguntas siguientes:

- ¿Fueron las actividades eficientes en cuanto a costos?
- ¿Se lograron los objetivos a tiempo?
- ¿Se implementó el programa o el proyecto en la forma más eficiente, comparada con las alternativas?

Impacto

Los cambios positivos y negativos producidos por una intervención para el desarrollo, directa o indirectamente, previstos o imprevistos. Esto incluye los principales impactos y los efectos resultantes de la actividad en los indicadores locales sociales, económicos, ambientales y otros indicadores de desarrollo. El examen debería tratar tanto los resultados previstos como los imprevistos, e incluir también el impacto positivo y negativo de factores externos, por ejemplo, cambios en los términos de comercio y condiciones financieras.

Al evaluar el impacto de un programa o un proyecto, es útil considerar las preguntas siguientes:

- ¿Qué sucedió como resultado del programa o el proyecto?
- ¿Qué diferencia real se obtuvo con la actividad realizada con respecto a los beneficiarios?
- ¿Cuántas personas se vieron afectadas?

Sostenibilidad

La sostenibilidad se refiere a medir si es probable que los beneficios de una actividad continúen después de haberse retirado la financiación de los donantes. Los proyectos deben ser sostenibles ambiental y financieramente.

Al evaluar la sostenibilidad de un programa o un proyecto, es útil considerar las preguntas siguientes:

- ¿Hasta qué punto continuaron los beneficios de un programa o un proyecto, después de cesar la financiación del donante?
- ¿Cuáles fueron los factores principales que influyeron en la obtención o no obtención de sostenibilidad del programa o el proyecto?

Parte III: Sistemas de evaluación y uso: herramienta de trabajo para revisiones y evaluaciones de pares

Este marco fue desarrollado en marzo de 2006, con base en un análisis completo de revisiones de pares efectuadas en un período de ocho años. Se diseñó para fortalecer la función de evaluación y promover la transparencia y la rendición de cuentas en las agencias para el desarrollo. Se desarrolló teniendo presentes las revisiones de pares y como mecanismo de administración para mejorar la práctica de la evaluación en las agencias de ayuda. Es una herramienta "viviente" diseñada para actualizarse en función de la experiencia.

1. Política de evaluación: papel, responsabilidad y objetivos de la unidad de evaluación
   - ¿Tiene el ministerio o la agencia de ayuda una política de evaluación?
   - ¿Describe la política el papel, la estructura de gobernabilidad y posición de la unidad de evaluación dentro de la estructura de ayuda institucional?
   - ¿Provee la función de evaluación una cobertura útil del programa total de cooperación para el desarrollo?
   - De acuerdo con la política, ¿cómo contribuye la evaluación al aprendizaje y la rendición de cuentas institucional?
   - ¿Cómo se conceptúa la relación entre la evaluación y la auditoría dentro de la agencia?
   - En países con dos o más agencias de ayuda, ¿cómo se definen y coordinan los papeles de las respectivas unidades de evaluación?

   ➤ *¿Se conoce e implementa adecuadamente la política de evaluación dentro de la agencia de ayuda?*
2. Imparcialidad, transparencia e independencia
   - ¿Hasta qué punto son independientes la unidad de evaluación y el proceso de evaluación de la administración de línea?
   - ¿Cuáles son los impulsores formales y reales que aseguran o restringen la independencia de la unidad de evaluación?
   - ¿Cuál es la experiencia de la unidad de evaluación en exponer los éxitos y fracasos de los programas de ayuda y su implementación?
   - ¿Es lo suficientemente transparente el proceso de evaluación para asegurar su credibilidad y legitimidad? ¿Se hacen consistentemente públicos los hallazgos de la evaluación?
   - ¿Cómo es el balance entre la independencia y la necesidad de interacción con la administración de línea que trata el sistema?

   ➤ *¿Perciben los actores no participantes en la evaluación los procesos y los informes de la evaluación como imparciales dentro y fuera de la agencia?*
3. Recursos y personal
   - ¿Tiene la evaluación soporte de recursos financieros y personal apropiados?
   - ¿La unidad de evaluación tiene asignado un presupuesto dedicado a ella? ¿Es anual o cubre varios años? ¿El presupuesto cubre actividades encaminadas a promover retroalimentación y uso de la evaluación y administración del conocimiento de la evaluación?
   - ¿El personal cuenta con pericia específica en evaluación y, si no, hay disponibilidad de programas de capacitación?
   - ¿Existe una política de contratación de consultores, en términos de calificación, imparcialidad y deontología?
4. Asociaciones de evaluación y generación de capacidad
   - ¿Hasta qué punto participan los beneficiarios en el proceso de evaluación?
   - ¿Hasta qué punto depende la agencia de evaluadores locales o, cuando no es posible, de evaluadores de terceras partes interesadas de países asociados?
   - ¿La agencia se encuentra vinculada a evaluaciones dirigidas por asociados?
   - ¿La unidad apoya programas de capacitación y generación de capacidad en los países asociados?

   ➤ *¿Cómo perciben los asociados, los beneficiarios o las ONG los procesos y productos de evaluación promovidos por la agencia o el país en términos de calidad, independencia, objetividad, utilidad y orientación de la asociación?*

5. Calidad
   - ¿Cómo asegura la unidad de evaluación la calidad de evaluación (incluyendo informes y procesos)?
   - ¿La agencia dispone de lineamientos para realizar evaluaciones y los utilizan las partes interesadas pertinentes?
   - ¿La agencia ha desarrollado o adoptado estándares o normas de referencia para evaluar y mejorar la calidad de sus informes de evaluación?

   ➤ *¿Cómo se percibe en toda la agencia la calidad de los productos y procesos de evaluación?*
6. Planeación, coordinación y armonización
   - ¿La agencia tiene un plan de evaluación de varios años que describa las evaluaciones futuras, según un cronograma definido?
   - ¿Cómo está desarrollado el plan de evaluación? ¿Quién, dentro de la agencia de ayuda, identifica las prioridades y cómo?
   - En los miembros del CAD donde la responsabilidad de la AOD es compartida entre dos o más agencias, ¿cómo está organizada la función de evaluación?
   - ¿La unidad de evaluación coordina sus actividades de evaluación con otros donantes?
   - ¿Cómo se coordinan las actividades de evaluación en el campo? ¿Está centralizada o descentralizada la autoridad para la evaluación?
   - ¿La unidad de evaluación está vinculada a evaluaciones conjuntas o de múltiples donantes?
   - ¿Hace uso la unidad de evaluación o la agencia de ayuda de la información de evaluación proveniente de otras organizaciones donantes?
   - ¿En qué forma evalúa la agencia la eficacia de sus contribuciones a las organizaciones multilaterales? ¿Hasta qué punto depende de los sistemas de evaluación de agencias multilaterales?
7. Difusión, retroalimentación, administración del conocimiento y aprendizaje
   - ¿Cómo se divulgan los hallazgos de la evaluación? Además de los informes, ¿se utilizan otras herramientas de comunicación (comunicados de prensa, conferencias de prensa, resúmenes, informes anuales con síntesis de los hallazgos)?
   - ¿Cuáles son los mecanismos vigentes para asegurar retroalimentación de los resultados de evaluación a autores de políticas, personal operativo y el público?
   - ¿Qué mecanismos vigentes existen para asegurar que el conocimiento de la evaluación sea accesible al personal y a las partes interesadas pertinentes?

   ➤ *¿El personal de la agencia considera la evaluación como una 'herramienta de aprendizaje'?*
8. Uso de la evaluación
   - ¿Quiénes son los principales usuarios de las evaluaciones dentro y fuera de la agencia de ayuda?
   - ¿Responde la evaluación a las necesidades de información expresadas por el Parlamento, la oficina de auditoría, el gobierno y el público?
   - ¿Existen sistemas en funcionamiento para asegurar un seguimiento y una implementación de los hallazgos y recomendaciones de la evaluación?
   - ¿Cómo promueve la agencia de ayuda o el ministerio el seguimiento de los hallazgos de las partes interesadas pertinentes (a través de, por ejemplo, grupos de dirección, paneles de asesores y juntas de sondeo)?

001El camino.indd 511 08/11/2010 03:28:02 p.m.

- ¿Se aseguran los vínculos con los procesos de toma de decisiones para promover el uso de la evaluación en la formulación de políticas?
- ¿Existen ejemplos recientes de cambios operativos y de política importantes activados por los hallazgos y recomendaciones de la evaluación?
- ¿Existen ejemplos de cómo sirve la evaluación como mecanismo de rendición de cuentas?

➤ *¿Cuáles son las percepciones de los actores no participantes en la evaluación (departamentos operativos y de políticas, oficinas de campo, etc.), con respecto a la utilidad y la influencia de la evaluación?*

Parte IV: Estándares de calidad de evaluación del CAD

Los CAD Evaluation Quality Standards fueron aprobados en 2006 para una aplicación de fase de prueba de tres años. La experiencia con el uso de los estándares por parte de los miembros y asociados interesados aportará la información para el acuerdo final de los estándares en 2009.

Los estándares soportan evaluaciones que se adhieren a los *CAD Principles for Evaluation of Development Assistance*, incluyendo imparcialidad, independencia, credibilidad y utilidad, y deberían leerse en conjunto con esos principios. Los principios se centran en los arreglos administrativos e institucionales de los sistemas de evaluación dentro de las agencias para el desarrollo; en contraste, los estándares proporcionan orientación sobre la realización de evaluaciones y para informes. Los estándares identifican los pilares básicos necesarios para un proceso y producto de evaluación de calidad, constituyen una guía de buenas prácticas y aspiran a mejorar la calidad de la evaluación para el desarrollo. Aunque no es obligatorio utilizar los estándares en cada evaluación, deben aplicarse lo más ampliamente posible y ofrecerse una breve explicación cuando esto no sea posible.

1. Fundamento, propósito y objetivos de una evaluación
   1.1 El fundamento de la evaluación
   Describe por qué y para quién se realiza la evaluación y por qué se realiza en un momento dado.
   1.2 Propósito de la evaluación
   El propósito de la evaluación está de acuerdo con la función de aprendizaje y la rendición de cuentas de las evaluaciones. Por ejemplo, el propósito de la evaluación puede ser:
   - Contribuir a mejorar una política, procedimiento o técnica de ayuda
   - Considerar una continuación o descontinuación de un proyecto o programa.
   - Contabilizar los gastos de ayuda para las partes interesadas y contribuyentes.

   1.3 Objetivos de la evaluación
   Especifican qué se intenta lograr con la evaluación.
   Por ejemplo:
   - Verificar los resultados (producto, resultado, impacto) y evaluar la eficacia, eficiencia y pertinencia de una intervención para el desarrollo específica.
   - Proporcionar hallazgos, conclusiones y recomendaciones con respecto a una política, programa, etc., específico.

2. Alcance de la evaluación
   2.1 Alcance de la evaluación
   El alcance de la evaluación se define claramente especificando los temas cubiertos, los fondos realmente gastados, el tiempo empleado, los tipos de intervenciones, la cobertura geográfica, los grupos objetivo y otros elementos de la intervención para el desarrollo tratados en la evaluación.
   2.2 Lógica de la intervención y hallazgos
   El informe de la evaluación describe brevemente y evalúa la lógica de la intervención, y diferencia los hallazgos a diferentes niveles: insumos, actividades, resultados e impactos.
   2.3 Criterios de evaluación
   El informe de evaluación aplica los cinco criterios del CAD para evaluar la ayuda para el desarrollo: pertinencia, eficiencia, eficacia, impacto y sostenibilidad. Los criterios aplicados para la evaluación dada se definen en términos inequívocos. Si no se aplica un criterio particular, esto se explica en el informe de la evaluación, al igual que los criterios adicionales aplicados.
   2.4 Preguntas de evaluación
   Las preguntas que se hicieron y cualquier modificación a las preguntas originales, se documentan en el informe para que los lectores puedan evaluar si el equipo de evaluación las ha evaluado lo suficiente.
3. Contexto
   3.1 El contexto de desarrollo y político
   El informe de evaluación presenta una descripción del contexto de política pertinente a la intervención para el desarrollo y los documentos, objetivos y estrategias de la agencia para el desarrollo y sus asociados. El contexto del desarrollo puede referirse a: la economía regional y nacional y los niveles de desarrollo. El contexto político puede referirse a: estrategias de reducción de la pobreza, igualdad de géneros, protección ambiental y derechos humanos.
   3.2 El contexto institucional
   El informe de evaluación ofrece una descripción del entorno institucional y la participación de las partes interesadas pertinentes a la intervención para el desarrollo, de manera que pueda identificarse y evaluarse su influencia.
   3.3 El contexto sociopolítico
   El informe de evaluación describe el contexto sociopolítico dentro del cual ocurren la intervención y su influencia en el resultado y en el impacto de la intervención para el desarrollo.
   3.4 Arreglos de implementación
   El informe de evaluación describe los arreglos organizativos establecidos para la implementación de la intervención para el desarrollo, incluidos los papeles de los donantes y los asociados.
4. Metodología de la evaluación
   4.1 Explicación de la metodología utilizada
   El informe de evaluación describe y explica el método y el proceso de la evaluación, y discute su validez y confiabilidad. Reconoce las restricciones encontradas y su impacto en la evaluación, incluido su impacto en la independencia de la evaluación. Detalla los métodos y técnicas utilizados en la recopilación y la información de los

datos y su procesamiento. Se justifican las elecciones y se explican las limitaciones y deficiencias.

4.2 Evaluación de resultados
Se especifican los métodos para la evaluación de los resultados. Deben tratarse la atribución y los factores de contribución o de confusión. Si se utilizan indicadores como base para la evaluación de los resultados éstos pueden ser Emopt (específicos, mensurables, obtenibles, pertinentes y atenidos al tiempo).

4.3 Consulta a partes interesadas aplicables
Se incluye a las partes interesadas pertinentes en el proceso de evaluación para identificar problemas y hacer aportes a la evaluación. Se consulta tanto a los donantes como a los asociados. El informe de evaluación indica las partes interesadas consultadas, los criterios para su selección y describe su participación. Si no se consultó a todas las partes interesadas, se describen los métodos y las razones para la selección de los consultados.

4.4 Muestreo
El informe de evaluación explica la selección de cualquier muestra. Se identifican limitaciones sobre la representatividad de la muestra de la evaluación.

4.5 Equipo de evaluación
Los equipos de la evaluación deben poseer una mezcla de destrezas en evaluación y conocimiento temático, estar equilibrados en cuanto a los géneros, e incluir profesionales de los países o las regiones interesadas.

5. Fuentes de información
   5.1 Transparencia de las fuentes de información
   El informe de evaluación describe las fuentes de información utilizadas (documentación, encuestados, literatura, etc.) con suficiente detalle, de manera que pueda evaluarse la suficiencia de la información. Se incluyen listas completas de entrevistados y documentos consultados, en cuanto esto no esté en conflicto con la privacidad y la confidencialidad de los participantes.
   5.2 Confiabilidad y precisión de las fuentes de información
   La evaluación efectúa un cruce de validación y evalúa críticamente las fuentes de información utilizadas y la validez de los datos, usando varios métodos y fuentes de información.
6. Independencia
   6.1 Independencia de los evaluadores frente a las partes interesadas
   El informe de evaluación indica el grado de independencia de los evaluadores de la política, la función de operaciones y administración del agente ordenante, los ejecutores y los beneficiarios. Los posibles conflictos de intereses se tratan abierta y honestamente.
   6.2 Proceso de evaluación libre y abierto
   El equipo de evaluación puede trabajar con libertad y sin interferencia. Se le asegura cooperación y acceso a toda la información pertinente. El informe de evaluación indica cualquier obstrucción que pueda haber afectado el proceso de evaluación.
7. Ética de la evaluación
   7.1 Realización de la evaluación en forma profesional y ética

El proceso de evaluación muestra sensibilidad con respecto a los géneros, creencias, maneras y costumbres de todas las partes interesadas, y se efectúa con integridad y honestidad. Se protegen los derechos y el bienestar de los participantes en la evaluación. Deben protegerse el anonimato y la confidencialidad de los informantes cuando lo soliciten o lo exija la ley.

7.2 Reconocimiento de desacuerdos dentro del equipo de evaluación
Los miembros del equipo de evaluación deben tener oportunidad de disociarse de juicios y recomendaciones particulares. Las diferencias de opinión no resueltas dentro del equipo deben reconocerse en el informe.

8. Aseguramiento de la calidad
   8.1 Incorporación de comentarios de las partes interesadas
   Las partes interesadas tienen la oportunidad de comentar sobre los hallazgos, conclusiones, recomendaciones y enseñanzas. El informe de evaluación refleja estos comentarios y reconoce cualquier desacuerdo sustantivo. En disputas sobre hechos verificables, los evaluadores deben investigar y modificar el borrador donde sea necesario. En caso de opinión o interpretación, los comentarios de las partes interesadas deben reproducirse literalmente, por ejemplo en un anexo, siempre que esto no esté en conflicto con los derechos y bienestar de los participantes.
   8.2 Control de calidad
   El control de calidad se ejerce durante todo el proceso de evaluación. Dependiendo del alcance y complejidad de la evaluación, el control de calidad se lleva a cabo internamente o por medio de un órgano externo, revisión de pares o grupo de referencia. Los controles de calidad se adhieren al principio de independencia del evaluador.
9. Pertinencia de los resultados de la evaluación
   9.1 Formulación de los hallazgos de la evaluación
   Los hallazgos de la evaluación son aplicables al objeto que se evalúa y al propósito de la evaluación. Los resultados deben seguirse claramente de las preguntas de la evaluación y del análisis de los datos, mostrando una línea clara de evidencia para apoyar las conclusiones. Se explican las discrepancias entre la implementación planeada y la real del objeto que se evalúa.
   9.2 La evaluación se implementa dentro del tiempo y el presupuesto asignados
   La evaluación se realiza y los resultados están disponibles de manera oportuna en relación con el propósito de la evaluación. Los cambios no previstos de tiempo y presupuesto se explican en el informe. Se explican las discrepancias entre la implementación planeada y los productos de la evaluación.
   9.3 Recomendaciones y enseñanzas
   Las recomendaciones y enseñanzas son aplicables, dirigidas a los usuarios previstos y procesables dentro de las responsabilidades de los usuarios. Las recomendaciones constituyen propuestas procesables y las enseñanzas generalizaciones de conclusiones aplicables a uso más amplio.
   9.4 Uso de la evaluación
   La evaluación requiere un reconocimiento explícito y una respuesta de la administración con respecto al seguimiento previsto a los resultados de la evaluación. La administración asegurará la difusión sistemática, el almacenamiento y administración del producto de la evaluación para asegurar un fácil acceso y maximizar los beneficios de los hallazgos de la evaluación.

10. Integridad
    - 10.1 En las conclusiones se responden las preguntas de la evaluación
      El informe de evaluación presenta las respuestas a todas las preguntas y las necesidades de información detalladas en el alcance de la evaluación. Cuando esto no sea posible se suministran razones y explicaciones.
    - 10.2 Claridad de análisis
      El análisis se estructura con un flujo lógico. Los datos y la información se presentan, analizan e interpretan en forma sistemática. Los hallazgos y conclusiones se identifican claramente y fluyen en forma lógica del análisis de los datos y la información. Los supuestos subyacentes se hacen explícitos y se tienen en cuenta.
    - 10.3 Distinción entre conclusiones, recomendaciones y enseñanzas aprendidas
      Los informes de evaluación deben distinguir claramente los hallazgos, las conclusiones y las recomendaciones. La evaluación presenta las conclusiones, las recomendaciones y las enseñanzas por separado y con una clara distinción lógica entre ellas. Las conclusiones se sustancian con los hallazgos y el análisis. Las recomendaciones y enseñanzas se siguen lógicamente de las conclusiones.
    - 10.4 Claridad y representatividad del resumen
      El informe de evaluación contiene un resumen ejecutivo que proporciona una visión general del proyecto, destacando las principales conclusiones, recomendaciones y enseñanzas aprendidas.

Parte V: Documentos de orientación

Como respuesta a la necesidad de una orientación más específica en ciertos campos de la evaluación para el desarrollo y con base en las experiencias de evaluación y las normas y estándares ya descritos, se desarrollaron varios documentos para dirigir la política y la práctica de la evaluación. Algunos de estos documentos de orientación se presentan a continuación.

Guidance on Evaluating Conflict Prevention and Peacebuilding Activities: Working Draft for Application Period
(OCDE, 2008)

Este documento introduce los retos y las buenas prácticas emergentes para evaluar la prevención de conflictos y las actividades constructoras de paz.

Con crecientes proporciones de recursos de ayuda, tiempo y energía dedicados a la prevención de conflictos y a la búsqueda de la paz, existe un mayor interés por saber lo que funciona bien y lo que no y por qué. Esta guía busca responder estas preguntas proporcionando dirección a quienes realizan evaluaciones de prevención de conflictos y proyectos, programas y políticas de búsqueda de la paz. Debería habilitar el aprendizaje sistemático, mejorar la rendición de cuentas y, a la postre, mejorar la eficacia del trabajo en busca de la paz.

Algunos de los mensajes clave emergentes de la guía incluyen:

- Los donantes deben promover la evaluación sistemática y de alta calidad de todo el trabajo de prevención de conflictos y de búsqueda de la paz, incluyendo el que realizan asociados en la implementación, como las ONG.

- Las evaluaciones deben facilitarse con un mejor diseño de los programas.
- Se necesitan una intervención y unas estrategias de políticas coherentes y coordinadas para poder progresar hacia la paz.
- Se requiere la clarificación de conceptos y definiciones de prevención de conflictos y búsqueda de la paz.

Guidance for Managing Joint Evaluations *(OCDE, 2006)*

Esta publicación proporciona una guía práctica para los administradores de evaluaciones conjuntas de programas para el desarrollo, con el fin de incrementar la eficacia del trabajo de las evaluaciones conjuntas. Usa las experiencias presentadas en "Joint Evaluations: Recent Experiences, Lessons Learned and Options for the Future" y en la anterior guía *Effective Practices in Conducting a Joint Multi-Donor Evaluation* (2000).

El foco de esta publicación no es la evaluación participativa con sus técnicas para introducir en el proceso las comunidades de las partes interesadas, sino las evaluaciones realizadas conjuntamente por dos o más agencias de cooperación para el desarrollo. Esos enfoques cooperativos, ya sean entre varios donantes, varios asociados o una combinación de ambos, son cada vez más útiles en un momento en que la comunidad internacional le da prioridad a la responsabilidad mutua por los resultados para el desarrollo, y a los enfoques conjuntos para el manejo de la ayuda.

Las evaluaciones conjuntas tienen potencial de beneficios para los asociados; por ejemplo:

- Desarrollo de capacidad mutua y aprendizaje entre los asociados.
- Generación de participación y propiedad.
- Compartir la carga de trabajo.
- Aumentar la legitimidad de los hallazgos.
- Reducir el número total de evaluaciones y los costos transaccionales totales para los países asociados.

No obstante, el trabajo conjunto puede generar también costos y problemas específicos que representen cargas para las agencias donantes. Por ejemplo, obtener consenso entre las agencias y mantener procesos de coordinación efectiva puede ser costoso y tomar mucho tiempo; los retrasos en la compleción de evaluaciones conjuntas complejas pueden afectar de manera adversa su oportunidad y pertinencia.

Evaluation Feedback for Effective Learning and Accountability *(OCDE, 2001)*

Esta publicación pone de relieve los distintos sistemas de retroalimentación, y hace un bosquejo de las áreas identificadas como más aplicables para mejorar la retroalimentación en las evaluaciones. También esboza las principales preocupaciones y problemas en la retroalimentación de las evaluaciones y los medios para abordarlos.

Un problema principal es el de transmitir los resultados de la evaluación a varias audiencias de dentro y fuera de las agencias para el desarrollo. Así, la retroalimentación y la comunicación de los resultados de la evaluación constituyen parte integral del ciclo de evaluación. La retroali-

mentación eficaz contribuye a mejorar las políticas, los programas y prácticas para el desarrollo, ofreciendo a los autores de políticas la información pertinente para tomar decisiones basadas en buena información. Las diferencias entre las agencias en sus antecedentes, estructuras y prioridades significan que éste no es un campo en donde sea apropiado un enfoque de planos. Además, es necesario adecuar los métodos de retroalimentación a distintas audiencias objetivo. Sin embargo, es posible identificar varios campos de acción a distintos niveles.

Las acciones sugeridas para mejorar la retroalimentación en las evaluaciones incluyen:

- Dar los pasos para comprender cómo se da el aprendizaje dentro y fuera de la organización, e identificar dónde ocurren embotellamientos.
- Evaluar cómo pueden mejorarse la pertinencia y la oportunidad de la retroalimentación, y tomar medidas para que esto suceda.
- Desarrollar un punto de vista más estratégico de cómo pueden adecuarse los métodos de retroalimentación a las necesidades de las distintas audiencias.
- Hacer mucho más esfuerzo para encontrar mejores formas de involucrar a los participantes de los países asociados en el trabajo de evaluación, incluyendo la retroalimentación de las enseñanzas de la evaluación.
- Tomar medidas para incrementar el espacio y los incentivos para el aprendizaje dentro de la organización (tanto de las evaluaciones como de otras fuentes).

Guidance on Evaluating Humanitarian Assistance in Complex Emergencies (OCDE, 1999)

Esta publicación va dirigida a quienes participan en la comisión, diseño y administración de evaluaciones de programas de ayuda humanitaria.

Históricamente, la ayuda humanitaria ha estado sujeta a procedimientos de evaluación menos rigurosos que la ayuda para el desarrollo. Como la participación de la AOD asignada a la ayuda humanitaria ha aumentado, y la conciencia de su complejidad también, la necesidad de desarrollar metodologías apropiadas para su evaluación se hizo cada vez más evidente. Esta guía complementa la de los *CAD Principles for Evaluation of Development Assistance*, resaltando los aspectos de la evaluación que requieren atención especial en el campo de la ayuda humanitaria.

Red de la OCDE CAD sobre la Evaluación del Desarrollo
EVALUACIÓN DE LA COOPERACIÓN PARA EL DESARROLLO
Sumario de normas y estándares básicos

Un componente clave de la misión de la CAD Network on Development Evaluation es desarrollar normas y estándares acordados internacionalmente para la evaluación para el desarrollo. Estas normas y estándares constituyen una información básica para la política y la práctica de la evaluación, y contribuyen a concertar los enfoques de acuerdo con los compromisos de la Declaración de París sobre la Eficacia de la Ayuda. El cuerpo de normas y estándares esta basado en la experiencia y evoluciona con el tiempo para adaptarse al entorno cambiante de la ayuda. Las normas y estándares sirven como punto de referencia internacional, guiando los esfuerzos para mejorar los resultados para el desarrollo mediante la evaluación de alta calidad.

Las normas y estándares aquí resumidos deben aplicarse juiciosamente y adaptarse de manera cuidadosa al propósito, el objeto y el contexto de cada evaluación. Como este documento sumario no es un manual de evaluación exhaustivo, se alienta a los lectores a consultar el texto completo disponible en el sitio web de la CAD Network on Development Evaluation:

www.oecd.org/dac/evaluationnetwork.

001El camino.indd 520 08/11/2010 03:28:02 p.m.

Matriz de diseño para el Programa de Capacitación Vocacional de Educación Secundaria de Baldonia Exterior

001El camino.indd 521 08/11/2010 03:28:02 p.m.

001El camino.indd 522

Matriz de diseño para el Programa de Capacitación Vocacional de Educación Secundaria de Baldonia Exterior, página 1

Tema principal de la evaluación: ¿Debe reautorizarse este programa? **Enfoque general** Evaluación de impacto cuasiexperimental

Preguntas	*Subpreguntas*	*Tipo de (sub) pregunta*	*Medidas o indicadores*	*Objetivo o estándar (si es normativo)*	*¿Datos básicos?*
1. ¿Qué servicios prestó el programa y a quién?	1.A.1. ¿En qué áreas de destreza vocacional se capacitó a los participantes?	Descriptivo	Áreas de destreza vocacional que ofreció el programa para capacitación.	NA	NA
	1.A.2. ¿Hubo cambios en la participación según área de destreza, a través del tiempo?	Descriptivo	Igual al anterior, según número de participantes cada año NA.	NA	Sí
	1.B.1. ¿Qué servicios de apoyo ofreció el programa?	Descriptivo	Servicios de apoyo ofrecidos por el programa (por ejemplo, alfabetización, consejería).	NA	NA
	1.B.2. ¿Qué proporción de participantes recibió servicios de apoyo?	Descriptivo	Número y porcentaje de capacitados que recibió cada tipo de servicios de apoyo.	NA	NA
	1.C. ¿Cuáles fueron las áreas más populares de certificación seleccionadas por los capacitados en el programa de capacitación vocacional?	Descriptivo	Número de capacitados según áreas de certificación.	NA	NA
	1.D. ¿Hasta qué punto ofreció certificación el programa en las áreas proyectadas como de alta demanda para los próximos 5 a 10 años?	Descriptivo	Lista de áreas vocacionales proyectadas como de alta demanda para los próximos 5 a 10 años.	NA	NA
2. ¿Hasta qué punto se puso en práctica la igualdad de géneros en los servicios?	2.A.1. ¿Participó en el programa un número igual de hombres y mujeres?	Normativo	Número y proporción de hombres/mujeres que recibió capacitación vocacional.	Documentos de autorización del programa que indican un 50% de meta de participación femenina.	NA
	2.A.2. ¿Se relaciona el recibo de servicios de apoyo con el género?	Normativo	Proporciones de recepción de servicios de apoyo según géneros.	Documentos de autorización del programa que indican meta de igualdad de géneros.	NA

08/11/2010 03:28:03 p.m.

001El camino.indd 523

Matriz de diseño, página 1 (continuación)

Diseño	*Fuentes de datos*	*Muestra*	*Instrumento de recopilación de datos*	*Análisis de datos*	*Comentarios*
1.A.1. *One-shot*.	Registros del programa (MIS).	Para cada uno de los últimos 5 años.	Documento No. 1 de recuperación de registros	Recuento de frecuencias. Análisis de contenido.	Las fuentes de datos deben concordar; anotar cualquier discrepancia y explicar.
	Director del programa.		Guía de entrevistas de los funcionarios del programa.		
1.A.2. Series de tiempo.	Igual al anterior.	Igual al anterior.	Igual al anterior.	Igual al anterior, por año.	Un gráfico sería bueno aquí.
1.B.1. *One-shot*.	Registros del programa (MIS).	Censo de los últimos 5 años.	Documento No. 2 de recuperación de registros	Lista.	Revisar duplicados como M. Smith y Mary Smith.
1.B.2. Igual al anterior.	Igual al anterior.	Igual al anterior.	Igual al anterior.	Recuento de frecuencias.	Anotar que los participantes pueden recibir más de un servicio de apoyo.
1.C. *One-shot*.	Registros del programa (MIS).	Censo de los últimos 5 años.	Documento No. 2 de recuperación de registros.	Recuento de frecuencias.	Gráfico.
1.D. Series de tiempo.	Informes anuales del Ministerio de Trabajo sobre proyecciones de trabajo a corto, mediano y largo plazo.	Informes para cada uno de los últimos 5 años.	Documento No. 3 de recuperación de registros	Análisis de tendencia y proyección para cada área de certificación ofrecida en los últimos 5 años.	Anotar cambios en tendencias y la sensibilidad del programa hacia ellos. Anotar si hubo áreas de crecimiento potencial en las que el programa no ofreció capacitación.
2.A.1. Series de tiempo.	Registros del programa (MIS).	Para cada uno de los últimos 5 años.	Documento No. 1 de recuperación de registros	Recuentos de frecuencias por géneros; presentación como gráfico de líneas de modo que se vea clara la tendencia de 5 años. Comparación con estándar.	Presentar el estándar con línea en negrita en el gráfico de líneas, de modo que se distinga fácilmente.
2.A.2. Igual al anterior.	Igual al anterior.	Igual al anterior.	Igual al anterior.	Igual al anterior.	Anotar si hubo cambios en el tiempo.

08/11/2010 03:28:03 p.m.

Matriz de diseño, página 2

Preguntas	*Subpreguntas*	*Tipo de (sub) pregunta*	*Medidas o indicadores*	*Objetivo o estándar (si es normativo)*	*¿Datos básicos?*
3. ¿Fue eficaz el programa?	**3.A.** ¿Hasta qué punto se cumplieron o superaron los objetivos anuales de colocación de empleos?	Normativo	Tasas de colocación de empleos, por año.	Sí. 80% de los que culminan el programa.	NA
	3.B. ¿Hasta qué punto se cumplieron o superaron los objetivos anuales de tasas salariales promedio en la colocación de empleos?	Normativo	Salarios de colocación de empleos para cada año.	Sí. US$2 por hora para los años 1 a 3 y US$3 por hora para los años 4 y 5.	NA
	3.C. ¿Hasta qué punto se colocó a los participantes en empleos concordantes con su área de certificación?	Descriptivo	Área de certificación y de colocación de empleo del capacitado.	Sólo estándar implícito; tratado como pregunta descriptiva.	NA
	3.D. ¿Cuál fue la tasa de abandono del programa?	Normativo	Número de ingresados al programa cada año y número de graduados cada año.	Los documentos del programa indican que no debería ser más de un 10%.	NA
4. ¿Fue eficiente en costos el programa?	**4.A.** ¿Fue razonable el costo por participante en el programa en relación con programas similares?	Descriptivo	Costo por capacitado colocado comparado con otros programas similares de capacitación.	Sólo estándar implícito; tratado como pregunta descriptiva.	NA
5. ¿Hasta qué punto constituyó un problema la rotación de instructores?	**5.A.** ¿Cuál fue la tasa de rotación de instructores?	Descriptivo	Tasa anual de rotación y para instructores.	Sin estándar fijado. El estándar implícito es que debería ser baja.	NA
	5.B. ¿Cuánto tiempo hubo puestos de instructor vacantes?	Descriptivo	Duración promedio y rango de puestos vacantes.	Sin estándar fijado. El estándar implícito es que debería ser baja.	NA
	5.C. ¿Se encontraron reemplazos para instructores con igual calificación?	Descriptivo	Años de experiencia en enseñanza. Años de trabajo en el área de certificación.	Sin estándar fijado. El estándar implícito es que debería ser comparable.	NA

001El camino.indd 524 08/11/2010 03:28:03 p.m.

001El camino.indd 525

Matriz de diseño, página 1 (continuación)

Diseño	*Fuentes de datos*	*Muestra*	*Instrumento de recopilación de datos*	*Análisis de datos*	*Comentarios*
3.A. Series de tiempo.	Registros de capacitados por año para cada uno de los 5 años (MIS).	Censo de colocados, por año.	Formulario No. 4 de recuperación de registros.	Comparación de estándar cada año y acumulado de 5 años.	Necesidad de validar información en los registros confirmando tasas de colocación y de inicio con empleadores para una muestra de capacitados. Recordar que probablemente será un problema para los 2 últimos años.
One-shot.	Registros de capacitados para los empleadores el último año (MIS). Registros de empleadores.	Muestra aleatoria.	Guía de entrevistas para empleadores. Forma No. 1 de registros de empleadores.	Concordancia de información del MIS con información de empleador.	
3.B. Series de tiempo.	Registros de capacitados por año para cada uno de los 5 años.	Censo de colocados, por año.	Formulario No. 4 de recuperación de registros.	Comparación de estándar cada año y acumulado de 5 años.	Necesidad de validar información en los registros confirmando tasas de colocación y de inicio con empleadores para una muestra de capacitados. Recordar que probablemente será un problema para los 2 últimos años.
One-shot.	Registros de capacitados para los empleadores los últimos 2 años (MIS). Registros de empleadores.	Muestra aleatoria.	Entrevistas de empleadores. Formulario de guía de registros de empleadores.	Concordancia de información del MIS con información de empleador.	
3.C. *One-shot*.	Registros de capacitados para los 2 últimos años con área de certificación (MIS).	Censo de colocados, por año.	Formulario No. 4 de recuperación de registros.	Recuento de frecuencias de concordancias por año y acumulado.	
	Empleadores que contrataron a capacitados en los 2 últimos años.	Muestra aleatoria.	Entrevistas de empleadores. Formulario de guía de registros de empleadores.		
3.D. Series de tiempo.	Registros de capacitados por año para cada uno de los 5 años (MIS).	Censo.	Formulario No. 4 de recuperación de registros.	Comparación de estándar cada año y acumulado.	
4. *One-shot*.	Funcionario financiero del programa.	Censo –los 5 años de estados financieros.	Entrevistas.	Costo por participante. Costo por participante colocado.	Esperar poder comparar costo por capacitado colocado con el de otros programas similares de capacitación.
	Tasas de colocación (*ver* 3A).				
	Estados financieros del programa.				
	Evaluaciones existentes de programas similares de capacitación.		Revisión de literatura.	Análisis de contenido.	
5.A. Series de tiempo.	Registros de empleo del programa.	Los 5 años.	Formulario No. 5 de recuperación de registros.	Recuentos de frecuencia por año, rango y promedio.	
5.B. Series de tiempo.	Registros financieros del programa.	Los 5 años.	Formulario No. 5 de recuperación de registros.	Recuentos de frecuencia, rango y promedio.	
5.C. Series de tiempo.	Registros de empleo del programa. c.v.s.	Los 5 años.	Formulario No. 6 de recuperación de registros.	Comparaciones de personal c.v.s. en 5 años	

08/11/2010 03:28:03 p.m.

001El camino.indd 526

Matriz de diseño, página 2

Preguntas	*Subpreguntas*	*Tipo de (sub) pregunta*	*Medidas o indicadores*	*Objetivo o estándar (si es normativo)*	*¿Datos básicos?*
6. ¿Hasta qué punto constituyeron una preocupación los capacitados que abandonaron?	**6.A.** ¿Cuáles fueron los números y porcentajes de hombres y mujeres que abandonaron el programa cada año?	Proporciones descriptivas, por género.	Número y proporción de hombres y mujeres que iniciaron el programa, por año y tasas de abandono por año.	Documentos de antecedentes indican que 10% es aceptable; implícito que sería la misma tasa para cada género.	NA
	6.B. ¿En qué áreas de capacitación abandonaron?	Descriptiva.	Igual al anterior, por área de certificación.	NA	NA
	6.C. ¿Cuáles fueron las razones habituales para abandonar el programa?	Descriptiva.	Razones de abandono más frecuentes.	NA	NA
	6.D. ¿Cuán preocupados estuvieron los funcionarios del programa por los abandonos?	Descriptiva.	Percepción de tasas de abandono. Opinión de si se actuó sobre el problema.	Ninguno especificado.	NA
7. ¿Hasta qué punto ganan más los capacitados colocados de lo que habrían ganado sin el programa?	**7.A.** ¿Cómo son las tasas de retención de puestos, aumento de salarios y de despido de los participantes colocados, comparadas con otros de características similares que abandonaron el programa?	De causa y efecto.	Retención de trabajo de participantes colocados, salario inicial, aumentos de salario y tasas de promoción y despido en dos años, comparados con: i) otros contratados por las empresas para posiciones similares en un período comparable; ii) ingresos antes del programa; iii) ingresos de los que abandonaron el programa.	Ninguno especificado.	Sí, salarios de colocación.
	7.B. ¿Qué opinión tienen los empleadores del desempeño de los empleados colocados, comparado con otros contratados de características similares que no recibieron el programa de capacitación?	De causa y efecto.	i) Probabilidad de contratar participantes del programa de no haber programa y ii) contratar más graduados del programa; iii) opiniones sobre el desempeño comparadas con otros.	Ninguno especificado.	Salarios de colocación iniciales; salarios precapacitación; si hubo.

08/11/2010 03:28:03 p.m.

001El camino.indd 527

Matriz de diseño, página 1; continuación

Diseño	*Fuentes de datos*	*Muestra*	*Instrumento de recopilación de datos*	*Análisis de datos*	*Comentarios*
6.A. Series de tiempo.	Registros de capacitados por año, para cada uno de los 5 años.	Censo por año.	Formulario No. 4 de recuperación de registros.	Distribución de frecuencias de abandonos por año; porcentaje de abandonos por año, por género y totales.	
6.B. Series de tiempo.	Registros de capacitados por año, para cada uno de los 5 años.	Censo por año.	Formulario No. 4 de recuperación de registros.	Tabulación de áreas de certificación, por frecuencia de abandonos del programa, anual y acumulada.	¿Podría ejecutarse prueba de significación chi-cuadrado?
6.C. *One-shot*.	Registros de capacitados para los últimos 2 años.	Censo de abandonos del programa para el período de dos años.	Encuesta de participantes anteriores.	Distribución de frecuencias de razones de abandono del programa: i) desde las perspectivas de los participantes y ii) según opinión de los funcionarios del programa de capacitación.	Triangular.
	Funcionarios del programa de capacitación.	Funcionarios superiores.	Funcionarios del programa.		
6.D. *One-shot*.	Registros de capacitados para los últimos 2 años.	Censo de abandonos del programa para dos años.	Guía de entrevistas.	Análisis de contenido y recuentos de frecuencias.	
	Funcionarios del programa de capacitación.		Guía de entrevistas de empleadores.		
7.A. Diseño cuasiexperimental –grupos no equivalentes.	Empleadores y registros de empleadores.	Censo para dos años anteriores.	Formulario No. 1 de registros de empleador. Encuesta de participantes anteriores.	Análisis de contenido/ frecuencias.	Sin grupo de comparación al inicio del programa; necesario utilizar variables de reemplazo.
	Abandonos del programa.		Entrevistas de empleadores.		
7.B. *One-shot*.	Empleadores y registros de empleadores.	Censo para dos años anteriores.	Formulario Guía No. 1 de registros de empleador.	Análisis de contenido/ frecuencias.	Anotar limitaciones de diseño de SyE; confidenciales, sin acceso.

Fuente: los autores.

08/11/2010 03:28:03 p.m.

001El camino.indd 528 08/11/2010 03:28:03 p.m.